(2)

AD-A216 219

A THEORY OF INFORMATION PRESENTATION FOR DISTRIBUTED DECISION MAKING

FINAL REPORT

DAVID NOBLE
JOSEPH TRUELOVE
CARLA GROSZ
DEBBORAH BOEHM-DAVIS

Engineering Research Associates
1595 Springhill Road
Vienna, Virginia 22180

OCTOBER 15, 1989

Sponsored by

COGNITIVE AND NEURAL
SCIENCES DIVISION
Office of Naval Research
Contract N00014-84-C-0484

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION unclassified	1b. RESTRICTIVE MARKINGS none
2a. SECURITY CLASSIFICATION AUTHORITY 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release Distribution unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Engineering Research Associates	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Engineering Research Assoc. 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research Cognitive and Neural Sciences Division
6c. ADDRESS (City, State, and ZIP Code) 1595 Springhill Rd. Vienna, Va. 22180	7b. ADDRESS (City, State, and ZIP Code) 800 North Quincy Street Arlington, Va. 22217
8a. NAME OF FUNDING / SPONSORING ORGANIZATION see 7a 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61153N PROJECT NO. NR649 TASK NO. D05 WORK UNIT ACCESSION NO. 411B005

11. TITLE (Include Security Classification)

A Theory of Information Presentation for Distributed Decision Making

12. PERSONAL AUTHOR(S)
David Noble, Carla Grosz, Joseph Truelove, Deborah Boehm-Davis

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 6/84 TO 9/89	October 15, 1989	108

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Coordination, distributed decision making, schema, decision aids, training, man-machine interface (KR)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

(tactical)
This research sought to identify and test theories of cognitive processes important to distributed decision making, and to define theory-based principles for improving information presentations able to support coordination among spatially distributed decision makers. This final report summarizes the major accomplishments. These include: development and test of a model for cognitive processes important to recognition-primed decision making, investigation of recognition processes in decision making by experienced people, generalization of situation assessment in team decision making, development of a theory-based "plan representation" chart to aid plan supervision and coordination in the Navy Battle Group, and preliminary evaluation of this chart at the Naval War College.

Keywords:

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL David Noble	22b. TELEPHONE (Include Area Code) (703) 734-8800 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

FINAL REPORT ON

A THEORY OF INFORMATION PRESENTATION FOR DISTRIBUTED DECISION MAKING

This is the final report on "A Theory of Information Presentation for Distributed Decision Making" contract N00014-84-0484 between the Office of Naval Research and Engineering Research Associates. This research was conducted as part of ONR's Tactical Distributed Decision Making program. The contract was initiated July 15, 1984 and ended October 14, 1989.

Objectives and accomplishments

The objective of this contract was to develop a theory of information presentation tailored to support coordination among Battle Group decision makers. This theory would relate basic cognitive processes used in decision making to properties of information presentations able to directly affect these processes. When fully developed the theory would be concrete and detailed enough to guide the design of decision aid displays.

In 1984 when this research began, there had been little or no work investigating the relationship between cognitive models and decision making, and there was little interaction between researchers investigating decision making and those examining basic cognitive processes. For example the 1984 review on behavioral decision theory and 1984 review on schema theory shared only a single reference[1]. Furthermore, most decision research examined the alternative evaluation phase of decision making rather than the earlier and often more critical problem definition and option generation phases. Finally, most decision aids, when based on any theory at all, relied on theories from economics or game theory rather than on psychology.

Under this contract, ERA has attempted to link cognitive models to decision processes, and to generate a psychology-based theory useful for decision aid design. Specific achievements include:

- Developing and testing a concept of "schema-based decision making" which emphasizes the role of recognition in decision making. ERA's second interim report described experiments which demonstrated a role of situation assessment in decision making. Although we were not aware of it at the time, other researchers were also starting to emphasize the role of recognition in decision making and to document its importance in decisions made by experienced decision makers. In September 1989 the Army Research Institute sponsored a workshop on "naturalistic decision making" which drew in part on the ERA research.

---

[1]The two reviews were "The Nature and Functions of Schemas" by W.F. Brewer and G.V. Nakamura in the Handbook of Social Cognition, Erlbaum, Hillsdale, N.J. with 130 references and "Judgment and Decision: Theory and Application" by Gordon F. Pitz and Natalie J. Sachs in the Annual Review of Psychology, vol 35, with 142 references. The shared reference was "Psychological Status of the Script Concept" by R. P. Abelson in American Psychology, vol. 36.

- Developing a detailed model of the cognitive processes and memory organization used in the situation recognition phase of decision making. Though this phase of decision making has been characterized as "intuitive," recognition actually entails a considerable amount of information processing. ERA developed and tested a model of this process, which is the basis for the information presentation principles set forth in the first attachment to this report.

- Developing a concept for situation assessment software based on the cognitive model. In 1986 and 1987 ERA received additional funding under this contract to examine adapting the cognitive model for situation assessment software. ERA's current situation assessment system, which is currently scheduled to be transitioned to two operational sites, is based on this adaptation.

- Examining situation assessment in team decision making. In well-trained teams individual team members can anticipate what others will do in various situations and adapt their own decisions accordingly. In doing this, team members may assess other team members' situation assessments. ERA extended situation assessment to include an assessment of another team member's situation evaluation and decision criteria, and documented people's assessments of their partners in a task entailing team decision making under uncertainty.

- Developing and evaluating a plan representation chart at the Naval War College. This chart, which represented the war game plans of War College students, is an example of a theory-based information presentation intended to support coordination within the Naval Battle Group. Its design was guided by the cognitive theory. By preparing charts for several groups of students, ERA showed that it was possible to apply the abstract theoretical information presentation principles to concrete practical cases. No controlled studies of the chart's contribution to coordination were conducted, but students and staff at the War College thought that it probably would improve plan supervision and coordination.

Because ERA's research was motivated by actual problems in tactical decision making, this research has enjoyed unusual success at transitioning results to more applied research and eventually to Navy products. Two transitions have already occurred. The first, mentioned previously, is the adaptation of the cognitive model to situation assessment software. The second is developing methods for helping people associate ambiguous reports with ship tracks.

Overview of attachments

There are four attachments to this report. The first summarizes the theory developed by ERA under this contract. It describes the decision making environment for tactical decision makers, emphasizing the importance of situation assessment in distributed military decision making. It then summarizes a cognitive model of the processes proposed to support situation assessment and reviews information presentation and training principles derived from the model. It concludes with an example of a theory-based information presentation, the plan representation chart developed by ERA with the assistance of staff at the Naval War College.

The theory described in this attachment is not complete, for a fully developed theory of information presentation must await a better understanding of cognitive processes and of the actual decision making behavior by different types of people engaged in different types of tasks. Nevertheless, even in its current formative state, the theory described in this report may be sufficient to guide decision aid design in some circumstances.

The remaining three attachments describe several of the experiments conducted to test and refine the theory. Two of these experiments were originally described in ERA interim technical reports and the third was described at the annual Distributed Tactical Decision Making program review. The three attachments revisit these experiments, reinterpreting the earlier results to reflect insights acquired later in the research.

The second attachment is the manuscript detailing the results of our situation assessment experiments. Because of recent data reported in the psychology literature, ERA has reevaluated the data we collected in 1985. The manuscript describes our recent analyses and the support for our current situation assessment model.

The third attachment describes evidence for recognition-primed decision making, emphasizing that recognition and outcome evaluation processes may intertwine in many decision processes.

The fourth attachment is an IEEE proceedings article which details the ERA investigations into the role of situation assessment in coordination. This article extends the concept of situation assessment to include an assessment of others' assessments. It examines what people do when their decision seems to depend on second guessing what other members of their team will do.

BIBLIOGRAPHY

Technical Reports

Noble, D., & Truelove, J. (1985). Schema-based theory of information processing for distributed decision making. Vienna, Va.: Engineering Research Associates. (NTIS No. ADA175156).

Noble, D., Boehm-Davis D., & Grosz, C. (1986). A schema-based model of information processing for situation assessment. Vienna, Va.: Engineering Research Associates. (NTIS No. ADA163150).

Noble, D., Mullen R. (1987). The Adaptation of Schema-based Information Processing Theory to Plan Recognition. Vienna, VA: Engineering Research Associates.

Noble, D., Boehm-Davis D., & Grosz, C. (1987). Rules, Schema, and Decision Making (ERA No. R-125-87). Vienna, Va.: Engineering Research Associates.

Noble, D., Truelove, J. (1988). Information presentations for distributed decision making: observations at the Naval War College. Vienna, VA: Engineering Research Associates.

Noble, D.(1989). Application of a Theory of Cognition to Situation Assessment. Vienna, VA: Engineering Research Associates.

Articles

Noble, D. (1989). Schema-based Knowledge Elicitation for Planning and Situation Assessment Aids. IEEE Transactions on Systems, Man, and Cybernetics, 19, 473-482.

Submitted Article

Noble, D., Boehm-Davis, D., & Grosz, Carla Extensions of Hintzman's Model to Meaningful Materials. Submitted to Journal of Experimental Psychology: Learning, Memory, and Cognition.

Book chapter

Noble, D. & Mullen R. (1988). Information Presentations for Distributed Decision Making. In S. Johnson & A. Levis (Eds.) Science of Command and Control: Coping with Uncertainty (pp 127-133). Washington, D.C.: AFCEA International Press.

Oral Presentations

Noble, D. (1985, June). A Theory to Guide the Design of Situation Assessment Aids for Decision Making. Paper presented at the 8th MIT/ONR workshop of C3 systems. Cambridge, MA.

Noble, D. (1985, October). Methodology for the Development of Decision Aids. Paper presented at the 3th Command and Control Decision Aiding Conference, Laurel, MD.

Noble, D. (1986, June). Development of Schema for Decision Making. Paper presented at the 9th MIT/ONR workshop of C3 systems. Monterey, CA.

Noble, D. & Mullen, R. (1987, June). Information Presentations for Distributed Decision Making. Paper presented at the 1987 Joint Director of Laboratories C2 Symposium. National War College, Fort McNair, Washington, D.C.

Noble, D. (June 1987). Template-based Data Fusion for Situation Assessment. Paper presented at the 1987 Tri-Service Data Fusion Symposium. Applied Research Laboratory, Laurel, Maryland.

Noble, D. & Truelove, J. (1987, November). Schema-based Decision Making. Paper presented at the 5th Command and Control Decision Aiding Conference, Kansas City, Missouri.

Noble, D. & Mullen, R. (1987, November). Information Presentations for Distributed Decision Making. Paper presented at the 5th Command and Control Decision Aiding Conference, Kansas City, Missouri.

Noble, D. & Truelove, J. (1988, February). A Chart for Plan Representation: Observations at the Naval War College. Paper presented at the 6th Command and Control Decision Aiding Conference, San Diego, CA.

Noble, D. (1989, November). Distributed Decision Making Under Uncertainty. In precedings of 1989 IEEE International Conference on Systems, Man, and Cybernetics, Cambridge, MA.

ATTACHMENT 1

A THEORY OF INFORMATION PRESENTATION FOR DISTRIBUTED DECISION MAKING

Engineering Research Associates
1595 Springhill Road
Vienna, Virginia 22180

OCTOBER 15, 1989

Sponsored by

COGNITIVE AND NEURAL
SCIENCES DIVISION
Office of Naval Research
Contract N00014-84-0484

A THEORY OF INFORMATION PRESENTATION FOR DISTRIBUTED DECISION MAKING

TABLE OF CONTENTS

LIST OF FIGURES

TABLE

1.0 AIDING MILITARY TACTICAL DECISION MAKING

The theory of information presentation connects models of basic cognitive processes to information display principles. This theory is intended to aid tactical decision makers by improving the quality of their situation assessments. This assessment includes estimates of hostile intent and future hostile actions. It also includes understanding the implications of the situation to successful attainment of mission objectives.

While targeted on military decision makers and distributed decision making, the theory is intended to apply to all contexts in which trained personnel make risky decisions in uncertain and time-stressed environments. The theory's emphasis on situation assessment reflects the importance of assessment in the Navy tactical environment.

1.1 The tactical environment

This environment is characterized by tactical uncertainties, time stress, high workload, high stakes, and constant situation changes. Information about the situation may be missing and ambiguous. This information may also be misleading, planted by an intelligent adversary to encourage ineffective tactical decisions.

The decision makers are experienced and well trained. Their decisions are guided by military doctrine, which specifies general types of actions appropriate in various kinds of situations, and also by specific war plans, which detail actions to be carried out for the specific situations that may be encountered during the planned mission. Despite the guidance provided by doctrine and the specific war plans, military decisions makers are expected exercise initiative in order to exploit unexpected opportunities or to minimize unanticipated risks.

Decision makers are organized hierarchically. In executing the war plan, they must coordinate both vertically with their superiors and subordinates and horizontally with peers. Effective coordination depends on all decision makers sharing a common understanding of the plan and a common interpretation of the situation.

1.2 The importance of recognition in distributed decision making

The theory of information presentation focuses almost entirely on recognition processes because situation recognition and interpretation is often the most critical decision making step in the environment described above. A decision making process which emphasizes recognition has been labeled "recognition-primed decision making" (Klein, 1989). In this mode of decision making, experienced people adapt basic courses of action that have worked well in similar types of situations. While recognition-primed decision making does not preclude projecting consequences of various alternatives, outcome projection plays a much smaller role in this mode of decision making than it does in traditional utility-based models of decision making.

Until recently there was very little discussion or study of recognition-primed decision making. There is now a developing literature suggesting the importance of recognition processes in military environments, in expert decision making in general, and in many basic judgmental processes.

Situation recognition is clearly central to decisions made in the execution of a military plan, for the plan is organized around the different types of actions to be taken in different kinds of situations. Situation recognition is also important to plan coordination, for when several decision makers are individually following the same plan, then successful coordination depends on each of the decision makers reaching the same conclusions about the current tactical circumstances.

Recognition processes have been emphasized in recent theories of decision making in command and control, and these processes have been widely observed in field studies. One influential command and control theory, the SHOR paradigm (Wohl et al, 1984), popularized the important of situation assessment in military decision making. The "H" in SHOR designates its second step, the generation and evaluation of situation hypotheses. This model did not, however, break with the outcome calculation tradition of decision making, for the "O" and "R" steps entail estimating the consequences of candidate alternatives. Gary Klein (Klein, 1989), however, found in his field studies that fire ground commanders, command and control personnel, and tank platoon leaders did not often formulate several alternative options. Rather, they considered only that action customarily applied in a given type of situation, considering additional options only if the customary option seemed inadequate. Lipshitz (1988) in analyzing Israeli military decisions also noted that most decisions relied more on situation recognition rather than evaluating the consequences of alternatives.

Researchers are also emphasizing the importance of situation recognition to decision making in complex non-military environments. Connolly and Wagner (1988) proposed a general decision cycles model which emphasizes a cyclic interplay between a decision maker's cognitive map of the situation and his goals. Pennington and Hastie (1986) in their study of decision making by jurors observed that in reaching a verdict jurors developed alternative situation models for different verdict categories. They selected the verdict corresponding to the situation model best able to account for the evidence. Chi, Feltovich, and Glaser (1981) investigated differences between novice and expert physics problem solvers. They showed that experts classified physics problems using underlying physics principles which relate to the solution method and then adapted that method for a particular problem. Chase and Simon (1973) found that chess experts can reconstruct the positions of chess pieces on a briefly observed chess board more accurately than novices can. Experts seem to remember basic types of chess situations, and reconstruct chess board positions by placing pieces to fit these remembered situations.

Controlled laboratory experiments have also shown the importance of recognition to decision making and judgment. In an experiment by Brooks (1987) subjects were given a rule for classifying cartoon-like creatures distinguished by different features. They were trained by viewing some examples of each category. In subsequent tests subjects classified new examples that were similar to the ones shown in training faster than new examples which were not similar to the training cases, even when applying the rule was nominally as easy for all examples The four cards problem described by Wason and Green (1984) also showed the importance of recognition rather than formal rules for making category judgments. In these experiments subjects performed much better applying a rule in concrete situations than in applying the same rule expressed abstractly in formal logic problem.

1.3 Aiding recognition pr. cesses

According to the recognition-primed paradigm, experienced decision makers often base their decisions on situation recog..ition. They identify options by adapting previously examined alternatives considered to work well in situations similar to the current tactical situat' n.

Aids which support recognition processes of decision making either will represent a physical situation or will represent a planned course of action. These aids should:

- help people recognize the applicability of a promising type of action in a particular situation.
- help people avoid applying an action which is not applicable in the situation.
- help subordinates apply the same criteria used by their commanders when evaluating the appropriateness of a proposed action, thereby reducing coordination errors along the chain of command.

The underlying idea of aiding recognition processes of decision making is simple. We wish to develop tactical information displays which help people in complex highly pressured environments to "see" a situation or problem as it might be viewed by more experienced people in less stressed environments. "Seeing" means noticing those aspects of a problem important for deciding what to do. Larkin and Simon (1987) describe the importance of such "seeing" by a chess example.

> "Consider, for example, a physical chessboard which we would represent as a set of squares, each with an (x,y) location and connections to adjacent squares. With each square is associated the name of any piece on it. Any person can "see" on what squares the pieces lie and locate adjacent or nearby squares. These inferences come from the primitive production rules that everyone has. But a chess expert may "see" things in the board not evident to the non-expert observer. For example, an important feature on a chess position is an open file: a sequence of squares that are vacant, running from the player's side of the board toward the opponent's side. In what sense is this seeing if everyone cannot see it?" (page 71)

According to our theory (which unlike Simon's is not based on a production rule model of human information processing), people identify promising problem solution methods by activating in memory processed feature lists for previously solved problems. These feature lists include many different kinds of features useful for identifying and evaluating problem solutions. In this chess example an open file is useful feature, for it may suggest actions able to exploit the opportunities or minimize the risks associated with an open file. A diagam of a chess board which makes explicit such features as "open files" may help novices notice the features used by experts, and may help experts in stressful environments consider the factors they would consider under normal conditions.

Larkin and Simon re-emphasized the importance of including the right features in presented information in their conclusion when they stated

> "...although every diagram supports some easy perceptual inferences, nothing ensures that these inferences must be useful in the problem-solving process. Failing to use these features is probably part of the reason why some diagrams seem not to help solvers, while other do provide significant help" (page 99).

That paper, which attempted to explain why diagrams may be more efficient for representing certain kinds of features than text is, did not address how to identify features that support the inferences useful in problem solving. This is, however, our objective, at least for the kinds of inferences important in recognition-primed decision making in challenging tactical environments.

The remainder of this paper describes the cognitive theory for situation assessment and the information presentation principles derived from this theory. It also describes an example of a theory-based information presentation, the plan representation chart developed to represent the war game plans of students at the Naval War College.

2.0 THE COGNITIVE FOUNDATION

The theory of the memory organization and information processing that we are proposing belongs in the general class of exemplar-based models (Medin and Shoben, 1988). It was originally motivated by the schema theories of Rumelhart (1980, 1984), but has evolved considerably over the past few years since its description in our 1985 and 1986 interim technical reports (Noble et al, 1985, Noble et al 1986). The research which influenced its evolution the most are the MINERVA 2 model of Hintzman (1986), the Wittlesea (1987) experiments testing an exemplar-based similarity model, the Kahneman and Miller norm theory (1986), and the Interactive Activation Model of Rumelhart and McClelland (1982). Our theory fits within the Parallel Distributed Processing paradigm (Rumelhart et al, 1986).

2.1 A theory of recognition-primed decision making and situation assessment

Key features of the theory are:

1. In recognition-primed decision making, people identify promising options by recalling from memory previously experienced problems.

2. Previously experienced problems are stored in memory as separate episodes. Each episode is a list of linked processed features.

3. These lists include four main types of features. These are features for problem objective, problem solution, environment conditions, and emotional state.

4. Features may be represented at multiple levels of abstraction.

5. Objective and surface features of a new problem activate those processed feature lists whose features are similar to those of the new problem. Activation depends on a feature-based similarity match.

6. A new problem can activate several features lists in parallel.

7. Processing is both top down and bottom up. In top down processing features in one part of the a memory-resident list create expectations about the characteristics of other features in the list, and cause a search of the external problem to determine whether those expectations are confirmed. In bottom up processing a feature in the external problem activates memory-resident feature lists whose features match those of the external problem.

8. Feature list activation increases whenever there is a match between the features in the feature list and the features of the external problem. Feature lists that are activated the most may deactivate feature lists that are not activated as much.

<u>Organization of Memory--Linked Processed Feature Lists</u>

An individual's knowledge is proposed to be organized as episodes. In general, these episodes correspond to different experiences. In the context of recognition-primed decision making, these episodes are previously solved problems.

Each episode is a network of linked processed features. The concept of a feature includes all things which may be associated with an instance or episode. The descriptor *processed* is joined to the term *feature* to emphasize that the features linked in memory as part of episodes are not simply objective observables that all people can see. Rather processed features are the result of interactions between what is already stored in memory and the new situation. In our theory the feature lists contain four types of features. These are: objective or purpose, external context (the observable environment), internal context (including the emotions of the person), and behavior and actions. Though a part of the theory, internal context features (emotions) will not be considered further in this discussion.

Each of the features in the list may be represented at multiple levels of abstraction. These abstractions may correspond to different levels in a "kind of" taxonomic hierarchy. For example, a pet dog may represented as a specific dog, a dog, and an animal. Feature abstractions can also represent functions or capabilities. Features at this level of abstraction indicate the *meanings* of features at more concrete abstraction levels.

Information Processing--Activation of Linked Feature Lists

Situations are recognized when the processed feature lists corresponding to those situations are *activated*, or in Rumelhart's (1980) terminology, *instantiated*. The features in activated feature lists are flagged as corresponding to features in the external environment. These features may also be refined or specialized, so that their characteristics correspond to the particular characteristics of the features in the environment.

Because feature lists for a previously solved problem include features which specify the problem's solution, activating the feature lists identifies a solution method. Our theory proposes that experts identify promising solutions to a problem this way.

Feature lists become activated when their characteristics match the characteristics of the environment sufficiently well. Processing involves comparing features in the external environment with features in stored feature lists and activating the most similar lists. For example, if the surface features of an object match the surface features in processed lists for chairs, then the perceived object will be classified as a chair. In activating lists, the process of matching features is not limited to the surface form of features. Features that match at the meaning level of abstraction can activate feature lists even if these features do not match at the surface level. This kind of processing has been suggested for categorization or classification tasks. We are proposing that it is used in all kinds of tasks.

This feature match process occurs without conscious awareness. What is generated by this similarity process, however, may enter conscious thought and influence behavior and further processing.

Activation does not occur all at once. Rather it results from a sequence of combined top-down and bottom-up feature matches and list activations. In the case of recognition for decision making, the process may begin by matching the objective of the external problem with the "objective" feature in feature lists representing previously experienced problems and activating those feature lists with matching objectives. These newly activated feature lists contain features which specify the solution method for the problem represented by the list and also contain environment features which specify the characteristics that the external problem should have in order for that solution method to work. The environment features initiate the top down processing. They (probably unconsciously) cause attention mechanisms to examine the external problem to determine whether it contains the specified features. If it does,then the feature list in further activated. If it does not, then the feature list activation is reduced.

This top down processing can use a person's world knowledge. This world knowledge is contained partly in the abstract representations of features in the processed feature lists for previously solved problems and partly in mental structures which specify the characteristics and capabilities of objects. Use of world knowledge enables an external problem whose objects match the objects of a previously experienced problem functionally but not physically to activate the feature list for that problem. In top down processing world knowledge may cause an evaluation of physical objects in a problem to determine whether they have particular capabilities. The barrier evaluation problem described later illustrates the use of general world knowledge in solving a new problem which matches some features of previously solved problems only at an abstract level of feature representation.

Bottom-up processing occurs at the same time as the top-down processing. Salient surface features will activate feature lists containing those features. Once activated, these feature lists initiate the top-down activation process. This process suggests that situations are most easily recognized when their corresponding feature lists share many features with the environment.

If the external environment matches several different processed feature lists, then each of these may be partially activated. When several lists are equally activated, then the situation is ambiguous, and may be interpreted in alternative ways. The theory assumes that each activated feature list reduces the activation of other lists. Thus, any feature list significantly more activated than other will deactivate these other lists.

According to this theory differences between novices and experts can be understood in terms of these processed feature lists. Experts have more feature lists for problems in their area of expertize than do novices, and their feature lists contain better solution methods and better abstract meaning features.

2.2 Barrier evaluation--an example of situation recognition

This example reviews an ERA experiment performed in 1985 (Noble et al, 1986). It uses our theory to explain how people evaluated barriers. It describes the processed feature lists and the information processing for this problem, and shows how the exemplar-based model enables people to use old examples and world knowledge to evaluate unusual or novel cases.

The exemplar-based theory's predictions of people's evaluations was excellent. Correlations between the subjects' evaluations predicted by the theory and their actual evaluations ranged from about .5 to .95 for the twenty subjects. Correlation between the predicted barrier evaluations and the average of the subjects' evaluations was about .98.

2.2.1 Experiment and data

In these experiments, subjects were trained by being shown ten different examples of barriers. Figure 2-1 is an example of the training picture showing a "perfect" barrier. Subjects were told that this barrier has an effectiveness rating of ten. They were also told the reasons for that rating, expressed in terms of relevant barrier features. Some of these features are "surface" (platforms are close together) and some are "meaning" (passage through the barrier is very difficult).

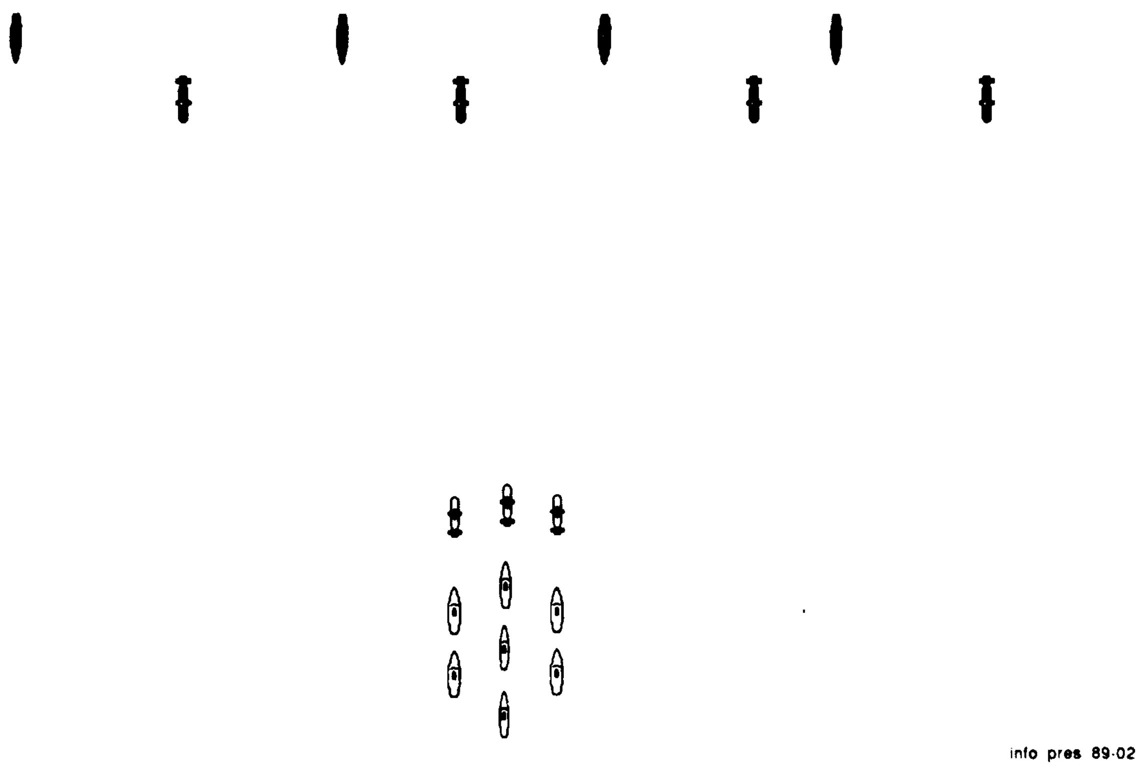


Figure 2-1. A barrier training picture. Subjects shown this picture were told "The barrier is both long and solid. The ships at the two ends are sufficiently far apart to make the barrier difficult to go around. The platforms are close enough together throughout its entire length to make passage through the barrier very difficult.

The other nine training pictures were similar to this one. Each picture depicted a different barrier composed of a row of ships and submarines. These barriers varied in the numbers and spacings of submarines and ships, but were otherwise the same. For each of these barriers subjects were given an effectiveness rating, expressed as a number between one and ten, and were told in what ways the barrier was strong or weak.

Although not mentioned to the subjects, the barrier effectiveness rating associated with each of the training pictures had been calculated using a formula that related two barrier features, length and size of largest gap, to barrier effectiveness. This formula, which is complicated, is:

$$\text{Barrier Effectiveness} = \min[G(\text{length}), H(\text{max gap})]^{.75} \times \max[G(\text{length}), H(\text{max gap})]^{.25}$$

where G and H are non-linear functions of length and maximum gap.

After training the subjects were asked to rate some new barriers. Each of these resembled the barriers seen in training, for each consisted entirely of a row of ships and submarines. The numbers and spacings of platforms in these barriers were new, however. Figure 2-2 is an example of one of these new barriers. The rating of this barrier, averaged over subjects, was 4.9.

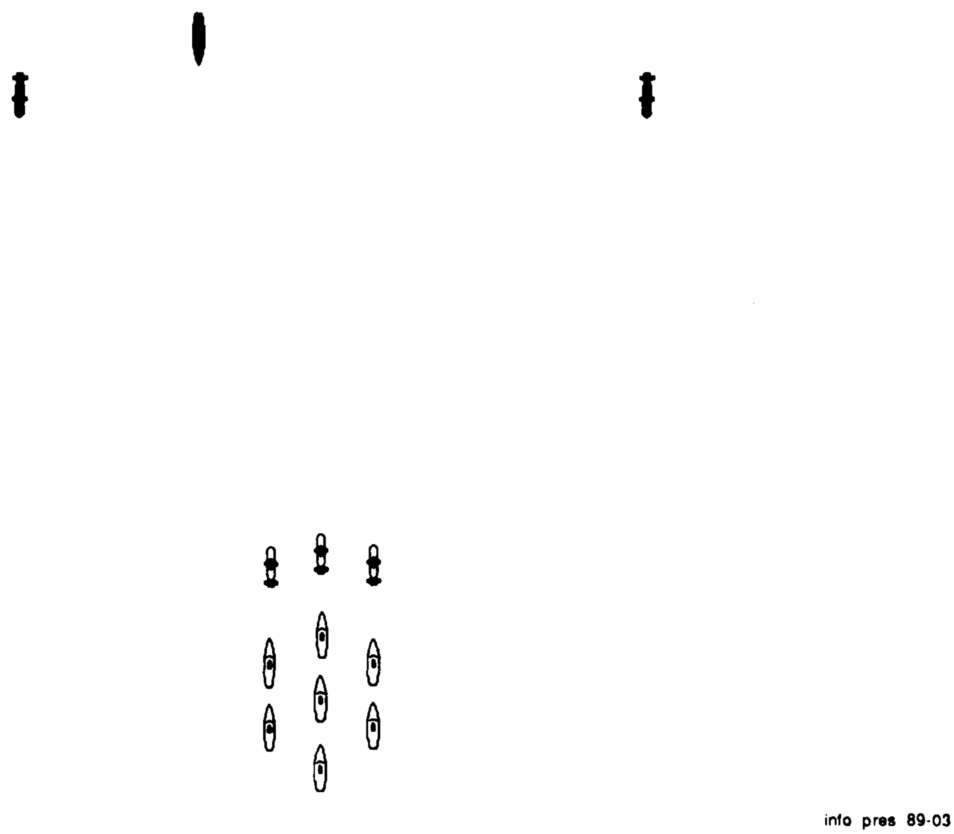


Figure 2-2. A barrier test picture. Average of subject effectiveness ratings was 4.9.

After rating ten barriers which generally resembled the training barriers, subjects evaluated another set of barriers. These barriers were not like any seen during training, for each had a feature not present in any of the previously presented pictures. Some of these barriers contained islands, some were adjacent to peninsula, and some were off center.

Although no special instructions concerning these features were provided, they strongly influenced the subjects' responses. For example, the barrier shown in Figure 2-3 has the same arrangement of ships and submarines as the barrier in Figure 2-2. The only difference is the addition of the island. This island increased the effectiveness rating from 4.5 for the barrier in Figure 2-2 to 7.0 for the example in Figure 2-3.

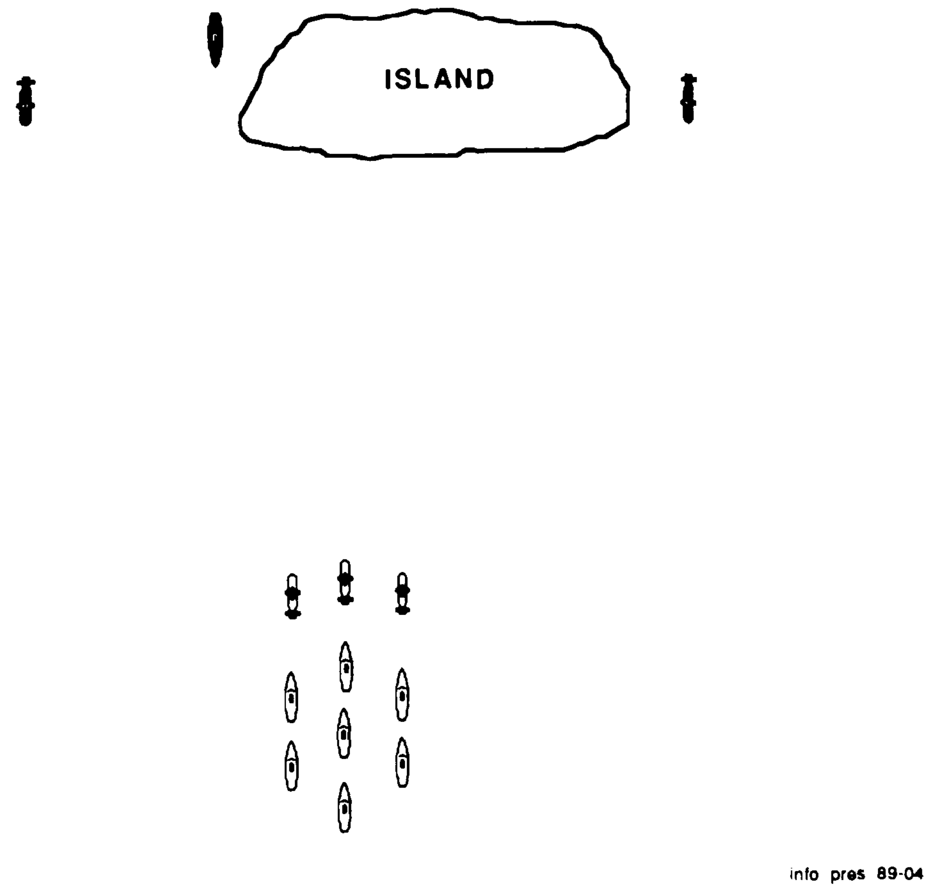


Figure 2-3. Another barrier test picture. The addition of the island in the barrier of Figure 2-3 changed the subject's effectiveness rating from 4.0 in Figure 2-3 to 7.0 in this figure.

2.2.2 Explanation: the importance of individual examples and the role of meaning features

Our theory proposes that during training the result of each barrier evaluation is stored in memory as a separate "solved problem". Evaluating a barrier can be regarded as "solving a barrier evaluation problem." These problems are encoded as processed feature lists with features for the objectives, environment, and problem solution method. Subjects evaluated new barriers by comparing in their minds the new barrier evaluation problem with the barrier problems seen in training, taking into account both the physical appearance of the barrier as well as its functionality.

Organization of knowledge in memory

The theory proposes that each problem presented during training is stored in memory as a processed feature list. Figure 2-4 illustrates this list for one of the training examples.

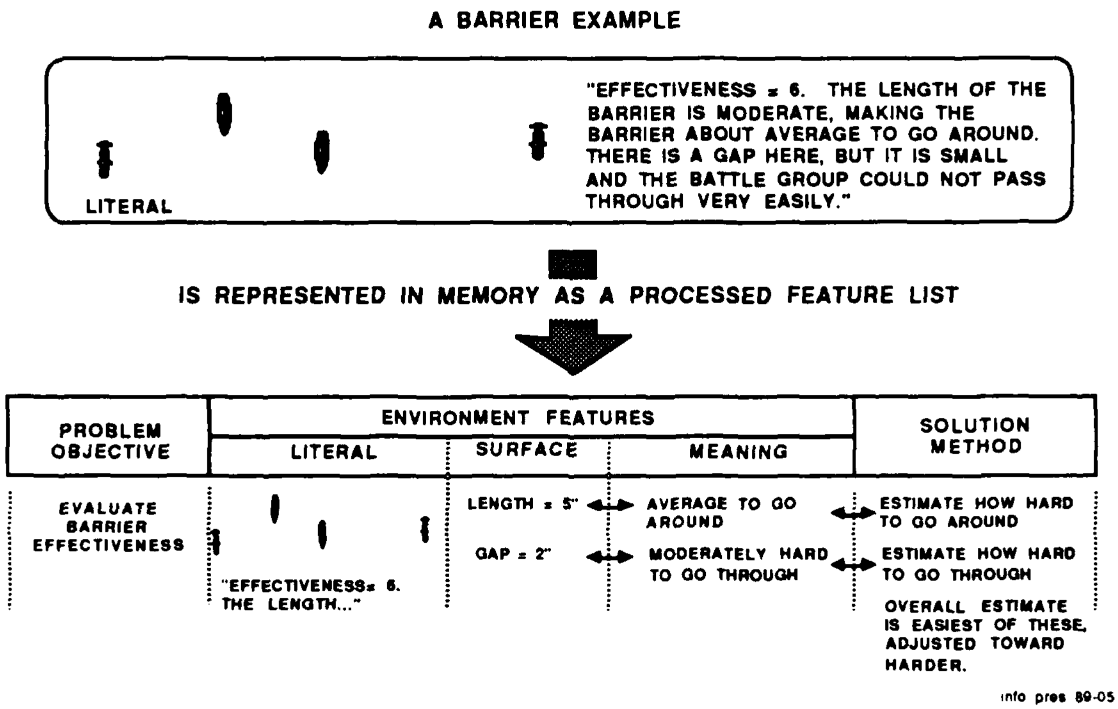


Figure 2-4. Representation of a barrier evaluation training problem as a processed feature list in memory.

The list contains several different types of features. These include:

1. Problem objective. The presence of problem objective in the feature list helps new problems to be related to old problems with similar objectives.

2. External environment features at several levels of abstraction.

   The literal representation is a problem image. It is included in the processed feature list to account for people's ability to recall and recognize surface characteristics of previously presented items, even if these surface characteristics are not relevant to the problem being solved.

   Surface features that are relevant to the problem solution. Surface features are countable or measurable objects or object relationships in the barrier. Relevant features are those useful for evaluating barrier effectiveness. These features are abstractions of the features in the literal representation. For these barriers, the relevant physical features are the barrier length and the size of the maximum internal gap.

Meaning "function" abstractions of these physical features. These features are abstractions of the surface features. They specify the reason why the surface features are relevant. Meaning features useful for barrier evaluation are the difficulty of going around the barrier and the difficulty of going through it.

3. General solution method. These are the general steps used to solve the problem. Steps to evaluate these barriers are: 1) estimate how hard it is to go around; 2) estimate how difficult it is to go through; 3) make an overall effectiveness estimate by combining these two estimates. Because a barrier is only as strong as its weakest link, this overall estimate is the easiest of these, adjusted toward the harder.

Note that it's the problem as a whole that is stored, not just the barrier itself. The problem representation not only includes a representation of the barrier, but it also includes the general solution method and problem objective. Furthermore, the features that are abstracted are those relevant to the solution method. Features not relevant to this method, such as the number of ships or submarines, may be stored as part of the literal representation, but are not stored in the list of abstracted surface and meaning features.

In this representation features at different levels of abstraction can be related to each other and to components of the general solution method. The arrows in Figure 2-4 indicate, for example, that the physical diagram length "5 inches" is related to the meaning feature "average to go around" and to the component of the general solution method "estimate how hard to go around."

Evaluation of Barriers

During training subjects learned to evaluate barriers using the general solution method based on how difficult it is to go around or to go through the barrier. The theory assumes that when asked to evaluate a new barrier, the subjects noticed that the new barrier resembles those that they had seen during training, and that therefore the evaluation method that worked for these previously seen barriers may work for the new one. Once they identify the general solution method, subjects estimate the difficulty of going around or passing through the barrier by comparing the characteristics of the new barrier with the characteristics of barriers encountered during training. Detailed steps in the evaluation procedure are:

1. The task of evaluating a new barrier activates previous tasks that presented similar problems (Figure 2-5). Physical features and task objective in the new problem activate old problems that share these features.

2. The activated old problems identify promising solution methods for the new problem, and also specify general properties (functional meaning features) which a problem should have in order for each of these solution methods to work. In this case, though the new problem activates many old problems, it activates only one solution method, the one shared by each of the activated old problems. In this method, barrier effectiveness is estimated by evaluating the two meaning features "how hard to go around" and "how hard to go through."

3. The two meaning features are evaluated by comparing the length and gap size of the new barrier with the length and gap size of each of the barriers seen during training.

4. If a surface feature in the new barrier resembles an evaluated previously seen feature, then the functional effectiveness for that surface feature can be estimated directly from that previously seen feature. Here the barrier length feature activates old barriers with similar lengths, and uses the length effectiveness rating of these activated "old" barriers to estimate the length effectiveness of the new barrier.

5. The length effectiveness estimate takes into account general world knowledge as well as the effectiveness estimates of the old barriers. For new barriers that resemble the old barriers closely in the length feature, this estimate is an extrapolation based on the old barrier. For example, if a barrier with a 5" diagram length were average to go around, a barrier with a 6" length would be judged harder than average to go around. The overall estimate of length effectiveness is an average of the length effectiveness estimates generated from the comparisons with each similar old barrier.

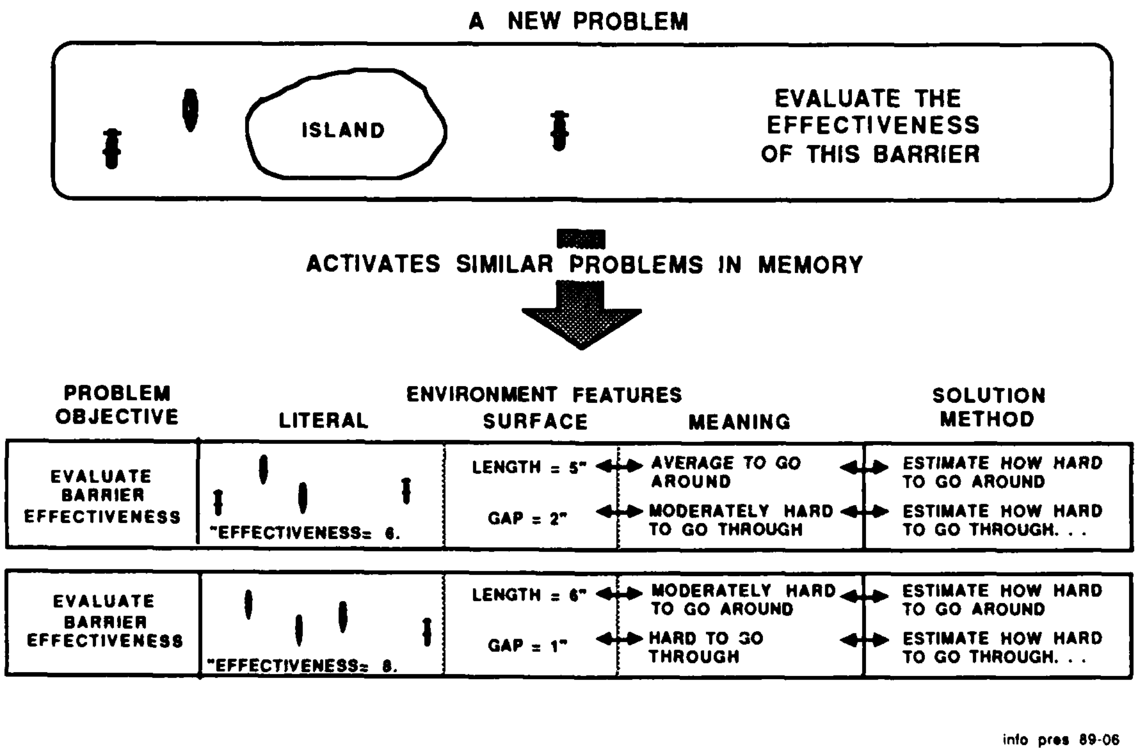


Figure 2-5. First step in barrier evaluation. A new barrier evaluation problem activates similar problems in memory.

The results of the evaluation at the conclusion of this step can be represented as an incomplete processed list like the one in Figure 2-6. This list contains the literal and problem objective information attained directly from the new problem statement. It also contains the information extracted so far from other similar barriers.

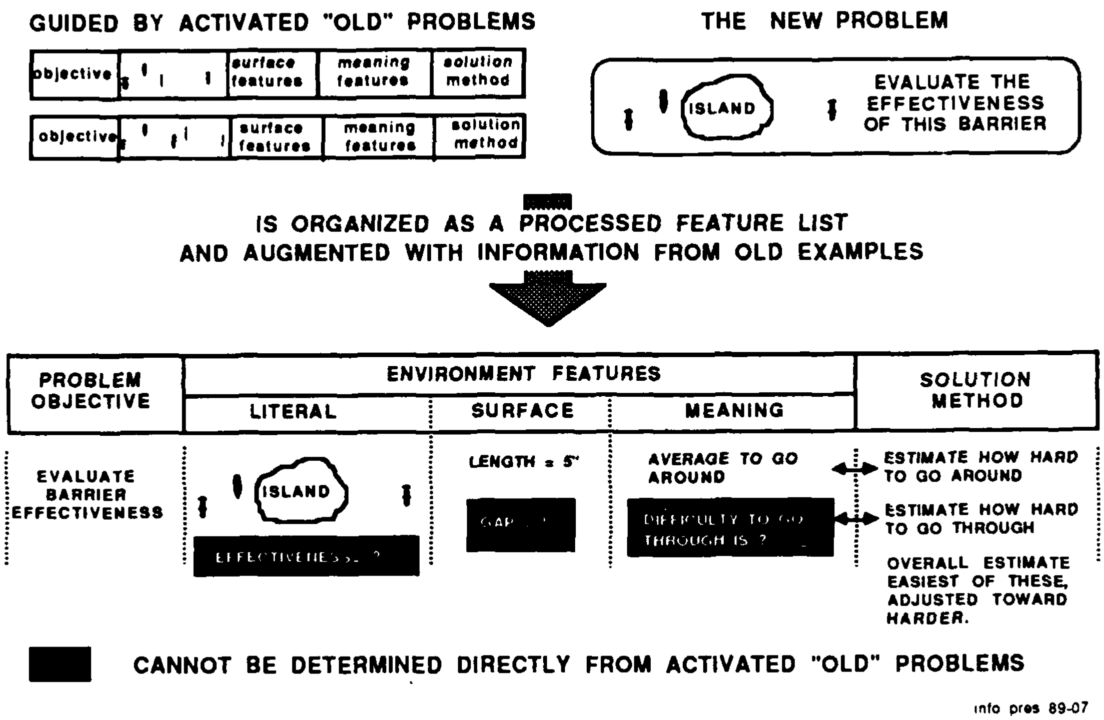


Figure 2-6. Second step in barrier evaluation. The new evaluation problem is encoded in memory as a partially processed feature list. This list specifies properties of the problem already observed or inferred, and indicates other properties that must be inferred for the barrier to be evaluated using the general solution method.

The feature list does not yet contain an estimate of the meaning feature "how hard to go through" because it unclear what size should be assigned to the gap. In addition, the feature list does not contain an estimate of overall effectiveness. These estimates cannot be attained directly from any of the old examples because the physical feature "gap size" in the new example does not match that feature in any of the old examples sufficiently well.

The subjects in our experiments were nevertheless able to estimate barrier effectiveness despite this feature mismatch, presumably because they had general world knowledge about the properties of ships and islands They appeared to use this world knowledge in the following way:

6. The meaning feature "how hard to go through" is estimated from general world knowledge about the functional properties of islands. The feature list specifies that gaps affect barrier effectiveness by determining how hard it is to go through the barrier. The fact that ships cannot sail through islands is retrieved from general world knowledge. Because ships cannot pass through islands, the barrier gap size is reduced by the length of the island. Therefore, for the barrier in Figure 2-6, the gap is reduced to a diagram size of 2", which in the old examples is moderately hard to go through.

The final step in the evaluation repeats on the barrier level the processing described earlier for estimating length effectiveness.

7. In this step (Figure 2-7) the values of the two meaning features, "how hard to go around" and "how hard to pass through," in the new problem are compared with the values of the functional features in the barriers of activated old problems. Each old barrier provides a separate estimate of barrier effectiveness. Since overall barrier effectiveness is estimated primarily from the weaker of these two features, barriers whose weaker features are close to those of the new barrier will have about the same effectiveness. The overall estimate is arrived at by taking the effectiveness of each individual old example and adjusting it, and then combining these data.

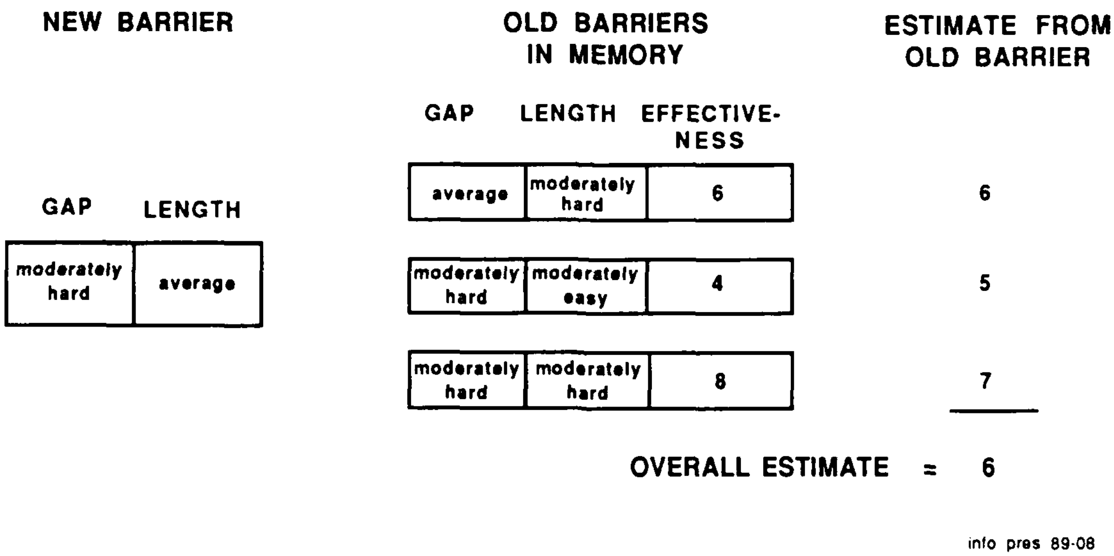


Figure 2-7. Final step in barrier evaluation. The new barrier is compared with each old one in memory. An independent estimate of the new barrier's effectiveness is derived from each of the old barriers by comparing the lengths and gap sizes of the new and old barriers. The estimate of the new barrier's effectiveness is the average of these independent estimates.

At the conclusion of this step, all of the features in the processed feature list created to evaluate the new barrier have been specified. This list can now be stored as an "old" example, and used for evaluating future barriers.

2.2.3 Summary: relationship to general theory

The explanation of how subjects evaluated barriers follows the general theory described previously. During training examples of barrier problems are stored in memory as separate processed features lists. A new barrier problem is solved by recognizing that the new barrier resembles these old problems. Recognition depends on activating the processed feature lists for similar old problems, a process facilitated by the similarity between the objectives and surface features of the old and new problems. The activated feature lists share a general solution method, and specify the meaning features that should be present in the new problem for this solution method to work. People determine whether these meaning features are actually present in the new problem by evaluating the surface features of the new problem. They can, when necessary, use general world knowledge to aid this process.

In this experiment the situation to be represented is a barrier evaluation problem, and not just the barrier itself. Therefore, the expectations of the situation representation are expectations about the problem rather than just expectations about the barrier.

The method for evaluating barriers illustrates a powerful general technique for interpreting novel situations: the interplay between bottom up data-driven object identification, top down guidance from the situation representation, and general knowledge. Subjects could recognize the objects in the barriers, the ships, submarines, and islands, using data-driven processing. They identified the general solution method and relevant functional properties of barriers top down from the situation representation. They evaluated the unusual barriers by accessing world knowledge to determine whether the objects recognized by data-driven processes have the functional properties specified top down by the situation representation.

3.0 THEORY-BASED TRAINING AND INFORMATION DISPLAY

This section describes general information presentation and training guidelines motivated by the cognitive model described in the previous section. Displays and training developed according to these guidelines are intended to exploit an understanding of the natural cognitive processes important to recognition-primed decision making. The training procedures should help people develop the memory content used by experienced decision makers. The information displays should facilitate access of this content.

These training procedures and displays should improve the quality of recognition-primed decisions by:

- helping people recognize the applicability of a promising type of action in a particular situation.
- helping people avoid actions which are not applicable in the situation.
- helping subordinates apply the same criteria used by their commanders when evaluating the appropriateness of a proposed action, thereby reducing coordination errors along the chain of command.

The first and second items above are related to two well known judgmental biases: the belief and and confirmation biases. The confirmation bias is the tendency to seek only information which confirms a current belief. The belief bias is the tendency to interpret available information so that it supports a current belief. Both biases can cause situations to be misinterpreted, which can lead to decision errors whenever decisions depend on situation assessment.

3.1 Information presentation principles

According to the cognitive model, during recognition-primed decision making the features of new problems activate processed feature lists which specify judgments and actions likely to be appropriate for solving the new problem. Aids which support recognition-primed decision making facilitate accessing those feature lists most useful for solving the target problem.

At the start of the recognition process the most salient features of a problem activate the feature lists of all previously solved problems sharing those features. When a previously solved problem is so activated, the rest of the features in its feature list, which were not directly activated by situation data, become indirectly activated. These indirectly activated features direct perceptual and inference mechanisms to search for corresponding features in the externally represented problem. Features that are found activate the processed feature lists containing corresponding features. Those feature lists that are most consistent with the external problem will become most strongly activated, and may suppress feature lists that are less consistent with the externally presented problem.

When this process is successful, it results in one or more activated processed feature lists, each of which corresponds to a solution method that worked in circumstances similar to the current ones. Each of these activated feature lists represents a promising alternative for solving the presented problem. These alternatives may then be evaluated further for suitability, and modified to suit particular circumstances. When no feature lists in memory are activated, then the decision maker may have difficulty identifying a promising approach to solving the problem.

This model can be used to predict how emphasizing certain kinds of features in an information presentation affects the number and types of processed feature lists activated, which in turn affects how well people recognize the applicability of a promising type of action in a particular situation, how well they avoid actions which are not applicable in the situation, and how well they avoid coordination errors along the chain of command. These predicted relationships are described below.

1. <u>Environment features</u> cue recall of similar situations for which particular problem solution methods were tried. The strength of the cue depends on how many and how closely the features in the presented information match the features in the processed feature list. Features that match several different processed feature lists will cue several different problem solution methods.

   Features that cue many different processed feature lists should help people consider alternative interpretations of a situation and to consider alternative solution methods. Emphasizing such features should help reduce the belief and confirmation biases, and thereby reduce selecting actions not applicable to the particular situation. For this reason, abstract representations of a feature found in many processed feature lists should decrease the belief and confirmation biases.

   More concrete representations that are found in fewer feature lists may cue fewer alternatives. These features, however, may often require less perceptual information processing in order to activate feature lists and thus may activate feature lists more strongly than do more abstract representations. Concrete feature representation should help people cue the usual solution methods in typical situations.

2. <u>Action features</u> depict the solution methods themselves. Concrete representations list the specific steps in the method, but may apply only when no unusual circumstances arise. More abstract representations may suggest the reason for the specific steps, and may provide guidance in unusual or unanticipated circumstances. These more abstract representations may, however, be more difficult to understand than concrete specific steps.

3. <u>Objective features</u> depict the objectives of the solution methods. They can be interpreted also as abstract action features, because they represent the steps to be completed at a general level. Emphasizing objectives activates feature lists of actions directed toward accomplishing those objectives. If there are several such feature lists, emphasizing these features reduces the belief and confirmation biases. If unexpected situations arise requiring that unplanned actions be taken, then emphasizing these features increases the chance that the action taken will be consistent with these objectives. If the objectives emphasized are those of the decision maker's supervisor, then emphasizing these features should reduce chain of command coordination errors.

4. Links between environment, action, and objective features help connect components of an action to the objectives of that component and to the environmental conditions which are indicators of that action. These links may help people to modify only those components of the action that need to be changed, while leaving undisturbed those components that remain useful.

5. Links between concrete and abstract variants of a feature help relate concrete representations of a feature to their more abstract meanings. These links may help concrete representations of a feature to activate the feature lists normally activated by more abstract representations. This should help people identify additional alternatives.

Table 3-1 summarizes the predicted decision consequences of emphasizing the different types of feature or feature relationships.

	Features or Relationships Emphasized	Impact on Decision Making
1.	Environment features	Help people identify promising alternatives.
2.	Concrete environment features	Help people recognize the usual solutions to typical problems.
3.	Abstract environment features	Help people recognize that a solution method can work for a new problem. Help people consider additional alternatives.
4.	Concrete action features	Remind people of standard way to solve a problem.
5.	Abstract action features	Enable people to solve a problem in more unusual circumstances.
6.	Objective features	Help people modify alternatives in ways consistent with objectives. May reduce chain of command coordination errors.
7.	Links between environment, action, and objective features	Help people identify specific components of a solution method that may need to be modified.
8.	Links between concrete and abstract variants of a feature	Help people identify additional alternatives by recognizing the significance of the abstract feature with respect to a solution method.

Table 3-1. Hypothesized impact of emphasizing various kinds of features or feature relationships on the effectiveness of the displayed information.

3.2 Development of training materials

The effectiveness of information presentations developed according to these principles stems from their ability to better access the processed feature lists most useful for solving a particular problem. Obviously, if these feature lists do not exist in memory, there is nothing for the presented information to activate, and consequently the presented information cannot support recognition-primed decision making. The purpose of the training principles described here is to put into memory the processed feature lists needed for recognition-primed decision making.

Training materials will support support recognition-primed decision making best if they can instill the processed feature lists of people who are expert rather than those who are novices in this problem area . The lists of experts presumably differ from those of novices because:

- experts, having solved more of a particular type of problem, have more processed feature lists for this type of problem. Consequently, the chances that a new problem will match one of these old feature lists is greater for experts than for novices.

- experts also have better features in their processed feature lists. The action features represent more general and more powerful solution approaches. The abstract meaning features associated with the components of the solution method cue the search for surface features critical to determining the applicability of a general problem solution method.

Because each of these lists represents a solved problem, they can be instilled by giving people different types of problems to solve. The theory suggests characteristics of training problems which may help develop and re-enforce the desired processed feature lists and therefore facilitate the training process. The desired feature lists may be instilled more efficiently if:

- training problems are organized about general solution methods. The training should emphasize that there are only a small number of basic different ways to solve a problem, and that a key step to solving a problem is identifying which of these basic ways is likely to work.

- each new problem should emphasize the abstract meaning features associated with the components of the general solution method. Training problems should make explicit the surface features associated with these abstract meaning features.

This theory predicts the following consequences from emphasizing general solution methods and different types of features in training:

- Emphasizing surface features associated with the most typical solution methods should help people solve standard problems in the standard way.

- Emphasizing meaning features associated with general solution methods will help people apply the correct solution method in less typical problems because these less typical problems may match on the meaning features even if they do not match on the surface features.

- If a new problem's surface or meaning features do not match the surface or meaning features of any old problem, then people will not recognize that a familiar general solution method will work for the new problem. Instead, they will have to develop the solution method from scratch.

Designing training materials requires identifying the basic different ways that experts solve problems, and then identifying those features most useful for cueing the most promising solution methods. These features are the ones that discriminate best among the types of situations for which different general solution methods will work. They can be identified by comparing the features elicited in the different problem classes. Those that occur in only a few classes may be better at cueing the right solution approach than those that occur in many.

This information must be attained from people experienced in solving these types of problems. Because people often have difficulty describing how they solve problems and why they chose to solve a problem in a particular way (Evans, 1988), people required to elicit expert knowledge have developed indirect methods for doing this. One of these methods, that used by Chi, Feltovich, and Glaser (1981) in their study of expert/novice differences, seems especially well suited for identifying the features which we propose are contained in the processed feature lists for solved problems.

In their work, Chi and her associates asked experts to classify physics mechanics problems by their method of solution, and then to list characteristics of problems associated with each of these methods. This led to a knowledge representation very similar to the one proposed in our model. The experts listed a small number of basically different ways such problems could be solved. They also listed the problem characteristics, (the surface and meaning features in our theory), useful for indicating whether any particular solution method may work for a particular physics mechanics problem.

4.0 THE PLAN REPRESENTATION CHART

The plan representation chart is an example of a theory-based information presentation. It was developed to demonstrate that the abstract presentation principles described in the last section could be applied in a concrete situation and that the resulting chart would have the intuitive appeal presumably inherent to information presentations developed in accordance with human memory organization and cognitive information processing.

These charts are described in detail in the 1988 ERA report "Information Presentations for Distributed Decision Making: Observations at the Naval War College" (Noble and Truelove, 1988). The following summary reviews the charts. It describes how the different features identified in Table 3-1 were used, and indicates their predicted impact on decision making. It also summarizes the results of the chart evaluations at the Naval War College.

4.1 Description of the chart

Overview

Plan representation charts summarize war plans. They are intended to help commanders determine whether a war plan still enables their forces to achieve their objectives, to help them identify those plan components that may need to be modified, and to help them avoid coordination errors when deciding on any modifications.

A plan specifies actions to be taken to attain mission goals. An overall plan may include several alternative plans. Each of these alternatives is associated with a particular set of plan assumptions which determine conditions under which it is to be executed. A plan representation chart reflects essential features of one of these alternative plans. Ideally, a separate chart would be prepared for each alternative plan developed in the planning process.

Figure 4-1 shows the overall organization of the chart. The chart is divided vertically into three main sections. The uppermost part specifies mission objectives. The middle part of the chart specifies plan assumptions, showing all relevant assumptions about possible enemy courses of action and environmental factors. The lowermost part of the chart depicts the directive. It shows the force organizational elements and the plan tasks assigned to each of these elements. These three sections of the chart correspond to the three general types of features--objective, environment, and action--in a processed feature list.

The plan chart is divided horizontally into two sections. The left section contains row labels. The right section depicts temporal relationships between designated tasks, plan assumptions, and plan objectives. Time increases along the horizontal axis in this part of the chart. Horizontal subdivisions in this right section represent different phases of the plan.

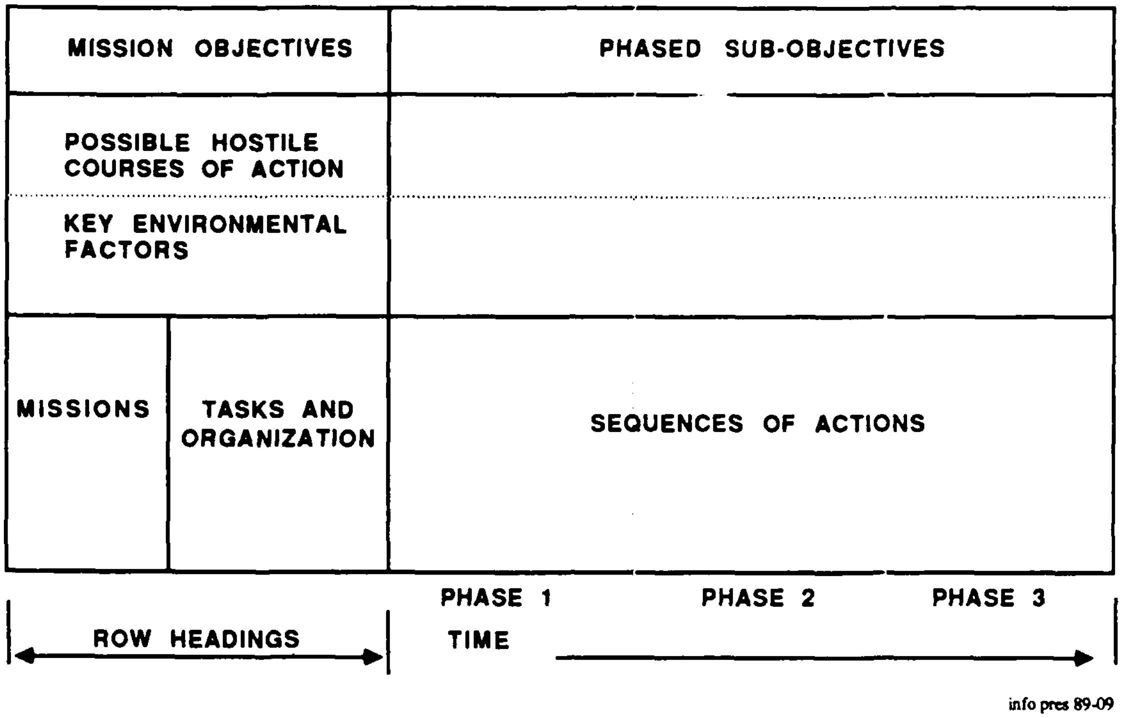


Figure 4-1. Overall organization of the plan representation chart.

Figure 4-2 shows one of the actual charts created to represent the plan of Seminar #7 in the War College Command and Control course.

Mission objectives section

Mission objectives are represented by sub-objectives selected to attain mission goals. In the chart, sub-objectives to be attained sequentially are displayed in series at the top of the chart. Sequentially addressed sub-objectives define the major phases of the mission. Sub-objectives to be attained simultaneously are drawn above and below each other.

The mission objectives and sub-objectives represent the planned actions at an abstract level. These planned actions are also represented at a much more concrete level in the directive section of the chart. The theory predicts that displaying objectives will reduce chain of command coordination errors.

Assumptions Section

Plan assumptions are those suppositions about events relevant to deciding at the time of plan execution which alternative plan to exercise. The assumptions section on each plan representation chart lists all suppositions relevant to selecting any of the alternative plans, and highlights those assumed to hold for the particular plan displayed on that plan representation chart. The assumptions section contains two types of assumptions: assumptions about possible Enemy Courses of Action and assumptions about the

		OBJECTIVES			PROTECT
					PROTECT MPS
				TG 70.7 BREAKOUT	
				DESTROY INTEL PLATFORMS	
ASSUMPTIONS	ECA	ATTACK 70.5/ ARG	SURFACE		
			AIR		AIR ATTACK
			SUBS		SUB ATTACK
			MINES		
		ATTACK TG 70.7	SURFACE		
			AIR		AI
			SUBS		SUE
			MINES	MINES	
		ATTACK MPS/ DIEGO GARCIA	KASHIN		
			ECHO II		
			BEAR A		
			MINES		
		INTERDICT SLRC/ URG	SUB		
			BEAR A		
			SAG		
	ENVIRON	NIGHT TIME			NIGHT
		SAG BEYOND 350 NMI OF CV		SAG BEYOND 350 NMI	
		SAG WITHIN TG STRIKE RANGE			SAG WITHIN
		SAG NEUTRALIZED			
		ARG BEYOND LRA RANGE			
DIRECTIVE		DESTROY INTEL PLTS	DGAR P3 (2)	LAUNCH 150800 — ATTACK AGI — ATTACK KASHIN	ASW/ ASUW PTRL/ ATTK
		PROTECT ARG	ARG	UNDERWAY 151200	
		TG 70.7 BREAKOUT	TG 70.7	EXIT HORMUZ	
		PROTECT MPS	MPS		DEPART DGAR 151800
			LEFTWICH	DETACH 150800 — TRANSIT TO MPS MODLOC	
		STRIKE KIEV SAG AND THE LAND FACILITIES	SENSORS	LAUNCH 150800 — LOCATE KIEV SAG — TRACK KIEV SAG	
			B-52		LAUNCH (2) H-3 151800
			CV STRIKE FORCE		LAUNCH H - 1.5 151930
			CV STRIKE SUPPORT GRP		

Figure 4-2 Plan representation chart prepared fo

TECT ARG / TG70.5

T MPS / DIEGO GARCIA

ATTACK SAG	ATTACK LAND FACILITIES
AIR SUBS	
HT TIME N TG STR RNG	SAG NEUTRALIZED ARG BEYOND LRA RANGE
	BEYOND LRA RANGE
RENDEVOUZ TG 70.5	
RENDEVOUZ DEPART FROM PAC	
ATTACK SAG 152100; RECOVER REFUEL REARM (H + 3) 160000	LAUNCH (4); ATTACK ADEN; RECOVER/ REFUEL/ REARM; LAUNCH (4)
ATTACK SAG 152100; RECOVER (H + 1.5) 152230	
	LAUNCH; SUPPORT ATTACK; RECOVER/ REFUEL/ REARM; LAUNCH

Task Group Commander of Seminar #7.

Environment. The outcome of tasks in earlier phases of the plan may be environmental assumptions of later phases. These assumptions indicate how the outcome of earlier phases of the plan affects the execution of later tasks.

The assumptions section corresponds to environmental features at a moderate level of abstraction. Assumptions which must hold for the plan depicted on the chart are highlighted on the chart. This highlighting indicates the properties which the tactical situation should have in order for the planned course of action to be appropriate. Assumptions which hold when alternate plans apply are listed on the chart, but are not highlighted. These are included in order to cue consideration of alternative options, and are intended to reduce the belief and confirmation biases.

During a military operation a computer-based display of these charts would strongly emphasize any assumptions indicated by tactical reports and intelligence estimates. This emphasis would prompt a commander to consider changing the plan if assumed conditions fail to occur or if conditions assumed by other plans arise.

Directive Section

The directive section of the chart specifies force organizational elements and the tasks and actions to be performed by these elements in order to achieve the mission objectives. This information is placed below the assumptions section, and occupies the lowest portion of the plan representation chart.

The left portion of the chart lists organizational elements responsible for the different tasks, and notes the general functions to be accomplished by these elements. These organizational elements are features of the planned action. The chart represents them at two levels of abstraction: a concrete level--the named platforms, and a more abstract level--the functions of these platforms.

The right portion of the chart portrays the time sequence of tasks and actions to be performed by these organizational elements. Each task or action is represented by a separate block on the chart. Blocks are arranged on the chart in the sequence that the tasks or actions are to be performed, and are associated with the appropriate operational phase.

Time increases from left to right on the chart, so that actions placed toward the right are expected to occur after those placed at the left. There is no explicit set time scale for the chart, and an inch on the chart may represent different time intervals at different points at the chart. Therefore, the precise starting time for a task cannot be inferred from the position of the task on the chart. Furthermore, since all blocks are approximately the same size, the duration of a task is not reflected by the length of the block representing the task.

The chart represents time in this non-literal way in order to accommodate the temporal uncertainties inherent to plans. A plan cannot specify the exact start times and durations of all tasks because some of these times cannot be predicted accurately when the plan is developed. It is not possible, for instance, to predict when hostile forces will choose to attack or when these forces will be detected.

The tasks in the directive section of the chart correspond to the action features at a fairly concrete level. The chart emphasizes both these tasks themselves, and also emphasizes the temporal relationship among these tasks. Because the duration of tasks and the time between tasks is not shown shown explicitly, these features are somewhat abstract.

Relationship between objectives, assumptions, and planned tasks

The vertical lines marking the different phases of a planned operation relate the objectives, assumptions, and planned tasks. Tasks to be performed in one phase support the objectives to be attained in that phase. Assumptions defined for each phase affect only the validity of the plan in that phase.

These vertical lines illustrate one way to implement the seventh and eighth items in Table 3-1. These lines relate concrete and abstract representations of the same feature (the tasks with sub-objectives) and relate action features with associated environment features.

The relationship between concrete and abstract representations of features was also shown within the objectives and directive sections of the chart. The general objective shown on the left was broken down into more concrete sub-objectives on the right. The concrete force organizational elements were also represented more abstractly as force element functions.

Showing these relationships should reduce the extent of modifications to plans which need to be changed, limiting task changes only to those affected by the no longer valid assumptions.

4.2 Chart evaluation summary

ERA developed the chart format with the help of the faculty and students at the Naval War College, with particular assistance from Mr. Frank Snyder, the War College faculty member responsible for teaching planning and decision making. Though guided by general principles of information presentation, the specific chart format and content evolved over a six month period.

The evaluation of the chart addressed several questions:

1. Would it be possible to represent the critical elements of actual war game plans in the chart?
2. Would the students who developed the plans easily understand the chart?
3. Could the chart promote a more uniform understanding of the plan among the planners?

4. Could the chart be dynamically updated to represent the progress of the plan toward achieving mission objectives?

5. Would information on the chart reduce the types of coordination errors observed during the war games?

6. Would the War College students think that the chart could improve Battle Group decision making?

The evaluation results were affirmative for each of these questions. ERA was able to develop charts for each of the war game groups involved in the evaluations. These charts were developed by encoding the text information in the two planning documents prepared by the students. These documents were the Commander's Estimate, which outlines the overall strategy and major assumptions of the plan, and the Directive, which specifies force organization, resources, and tasks.

To determine whether the charts were easy to understand and whether they could contribute to a common understanding of the plan, ERA asked several of the planners individually whether the chart correctly represented their war game plans. These students had no difficulty responding, indicating that they could easily understand the charts. Each of the planners thought that the chart did capture the key elements of the plan, but each also added an additional detail. Since each of the different planners added a different change, it seems likely that the charts can be used to reduce differences in plan understanding among the planners.

During the war games themselves, the war game commander reviews the progress of the plan to determine whether the plan still enables mission objectives to be achieved, and if not, how the plan should be changed. To test whether the charts could support this process ERA attempted to update the charts to track the progress of the plan. Though the manual process was awkward, chart updating was possible. In this initial interaction with the War College, the charts were displayed in the war game spaces, but the students were not asked to try to use them and did not use them. When asked about the potential value of the charts, most of the students felt that this type of information could improve Battle Group decision making if the charts were integrated into the computer-based Battle Group information systems.

Several coordination errors were observed during the games. Coordination errors among peer decision makers such as the Anti-air Warfare Commander and the Anti-Surface Ship Warfare Commander were not observed, probably because the plans were designed to minimize interactions among peer commanders. The errors that were observed were caused by misunderstandings hierarchically in the organization, between the Officer in Tactical Command and a warfare area commander. Because the misunderstandings were addressed by specific items on the charts, it seem plausible that these charts may reduce coordination errors in the Battle Group.

REFERENCES

Brooks, L. R. (1987). Decentralized control of categorization: the role of prior processing episodes. In U. Neisser (Ed.), Concepts and conceptual development: ecological and intellectual bases of categories (pp 141-174). New York: Cambridge University Press.

Chase, W. G., & Simon, H. A. (1973) The mind's eye in chess. In W. G. Chase (Ed.) Visual Information Processing (pp215-281). New York: Academic Press.

Chi, M. T. H., Feltovich, P. J. , & Glaser, R. (1981) Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Connolly, T. & Wagner, W. G. (1988) Decision Cycles. Advances in Information Processing in Organizations, 3, 183-205.

Evans, J. St . B. T.. (1988) "The knowledge elicitation problem: a psychological perspective. Behavior and Information Technology, 7(2), 111-130.

Hintzman, D. L. (1986) "Schema abstraction" in a multiple-trace memory model. Psychological Review, 93(4), 411-42.

Kahneman, D. & Miller T. (1986) Norm Theory: Comparing reality to its alternatives. Psychology Review, 93, 136-151.

Klein, G.A. (1989) Recognition-primed decisions. In W. Rouse (Ed.), Advances in Man-Machine Systems Research, 5, 47-92. Greenwich, CT: JAI Press, Inc.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. Cognitive Science, 11, 65-99.

Lipshitz, R. (1988) Making Sense of Decision Making: The Implausibility of Real World Decisions as Consequential Choice. Boston University, Center for Applied Behavior Science.

Medin, Douglas L., & Shoben, Edward J. (1988). Context and Structure in Conceptual Combination. Cognitive Psychology, 20, 158-190.

Noble, D., & Truelove, J. (1985). Schema-based theory of information processing for distributed decision making. Vienna, Va.: Engineering Research Associates. (NTIS No. ADA175156).

Noble, D., Boehm-Davis, D., & Grosz,C. G. (1986) A schema-based model of information processing for situation assessment. Vienna, VA: Engineering Research Associates. (NTIS No. ADA163150)

Noble, D., Truelove, J. (1988) Information presentations for distributed decision making: observations at the Naval War College. Vienna, VA: Engineering Research Associates.

Pennington, N. & Hastie, R. (1986) Evidence evaluation in complex decision making. Journal of Personality and Social Psychology, 51, 242-258.

Rumelhart, D. E. (1980) Schemata, the building blocks of cognition. In R.J. Spiro, B.C. Bruce, and W.F. Brewer (Eds.) Theoretical issues in reading comprehension . Hillsdale, N.J.: Erlbaum.

Rumelhart, D. E. (1984) Understanding understanding. In J. Flood (Ed.) Understanding reading comprehension (pp1-20). Newark, Delaware: International Reading Association.

Rumelhart, D. E. & McClelland, J.L. (1982) An interactivation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extension of the model. Psychological Review, 89, 60-94.

Rumelhart, David E., James L. McClelland & the PDP Research Group. (1986). Parallel Distributed Processing. Explorations in the Microstructure of Cognition. The MIT Press. Cambridge, Mass.

Smith, Edward E. & Medin, Douglas L. 1981. Categories and Concepts. Harvard University Press. Cambridge, Mass.

Wason, P. C., & Green, D. W. (1984) Reasoning and mental representation. The Quarterly Journal of Experimental Psychology, 36A, 597-610.

Whittlesea, W. A. (1987) Preservation of specific experiences in the representation of general knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13(1), 3-17.

Whittlesea, B. W. A., & Cantwell, A. L. (1987). Enduring influence of the purpose of experiences: Encoding-retrieval interactions in word and pseudoword perception. Memory & Cognition, 15(6), 465-472.

Wohl, J. G., Entin, E. E., Kleinman, D.L. & Pattipati, K. (1984) Human Decision Processes in Military Command and Control. Advances in Man-Machine Systems Research, 1, 261-307.

ATTACHMENT 2

EXTENSIONS OF HINTZMAN'S MODEL TO MEANINGFUL MATERIALS

Engineering Research Associates
1595 Springhill Road
Vienna, Virginia 22180

OCTOBER 15, 1989

Sponsored by

COGNITIVE AND NEURAL
SCIENCES DIVISION
Office of Naval Research
Contract N00014-84-0484

Abstract

This research tests the adequacy of the Hintzman (1986) model to explain how subjects use old examples to evaluate new instances and evaluates two extensions that increase its power when the materials are meaningful. In the first extension, people adjust feature values to reflect the relationship between feature characteristics and overall characteristics of the example. In the second extension, people replace a new feature with a functionally equivalent familiar feature. In the experiments, subjects were trained on examples of stylized "all-out attacks" and "barriers". For each example, they were shown a picture of attacking forces and were told the force's "effectiveness". Later they were asked to estimate the effectiveness of forces in similar dispositions. Both the original Hintzman model and the first extension to Hintzman's model adequately account for the data when the new instances contain only features seen in the training instances, although the first extension to the Hintzman model does seem to provide a slightly better account. When the test instances contain new types of features, neither the original Hintzman model nore the first extension can account for subjects' responses. The second extension can account for some, but not all, of the subjects' evaluations.

Extensions of Hintzman's Model to Meaningful Materials

In 1986, Hintzman proposed an exemplar-based model of schema that did not rely on the notion of an abstracted, centralized prototype. In that paper, Hintzman demonstrated that his model could account for some properties of memory (such as the finding that prototypes are more easily recognized than individual previously-seen exemplars) when drawing solely from a collection of examples. Whittlesea (1987) has shown that this type of model can also be used to explain human performance using nonsense/meaningless materials comprised of 5-letter pseudowords. In this study, Whittlesea found that categorization performance depended on similarity to previously-seen instances rather than on similarity to a prototype, once typicality effects were unconfounded from similarity. Kahneman and Miller (1986) also found that norms for meaningful materials appear to be constructed from information about specific instances rather than from precomputed expectations.

Our objective in this research was to determine how well Hintzman's (1986) model can be used within the domain of simple meaningful materials, and how easily it can be extended to accomodate more complex examples. For this research, we chose "military" environments. These military situations would not appear realistic to military planners, but did capture general knowledge about situations that our subjects (university undergraduates) would be familiar with. In the first experiment, the situations to be evaluated were "all-out attacks". In the second experiment, the situations were "barriers".

These situations were represented in drawings where some number of hostile forces (which varied) confronted a battle group. In this study, subjects were initially shown examples of the situations. Each had a rating of the effectiveness of the hostile force in the shown instance and had a feature-based explanation of the rating. Figure 1 shows an example of an all-out attack training picture with its associated rating. Subjects were then shown new instances and asked to evaluate how good these new instances were as examples of all-out attacks.

Insert Figure 1 about here

Overview of Models

Hintzman Model, as adapted in this experiment

Following Hintzman (1986), it is proposed that subjects might generate the rating for a new instance as a weighted average of the effectiveness ratings of similar exemplars (see Figure 2). The weights would be computed from a similarity measure between the new example and each of the instances stored in memory. In this model, episodes are experienced and encoded as entities and concept formation occurs on the basis of those stored instances. Each of the entities is encoded as a combination of primitive properties which are either present or absent. When a new example is encountered, it is compared with each stored instance and a similarity measure is computed based on common primitive properties. This similarity measure determines the importance that the actual primitive property values of the old instance will have in the response that is generated. The response is derived as the weighted average of the episodes to which it was similar.

Thus, one could predict that when participants are shown a test item that has not been seen before and are asked to rate the test item, the examples stored in memory would be recruited and the predicted rating for the new item would be calculated as a weighted average of the ratings of the items recruited. The ratings of the stored examples that are more similar to the test item are weighted more heavily than the ratings of the examples that are less similar to the test item.

There is evidence to suggest that people may process material in this way. Whittlesea (1987), using non-meaningful materials, showed that similarity to training instances was more important than similarity to an abstracted prototype. In his study, he constructed CVCVC letter combinations that were variants of two prototypical categories (FURIG and NOBAL). Performance, as measured by the gain in perceptibility of new instances of the category, was increased when the new items were similar to the original items, not when the new items were similar to the prototypes. Brooks (1987) reviews a number of studies which suggest that recognition memory, perceptual identification, and classification performance are all sensitive to the actual instances seen and to the processing performed on those instances.

There is also evidence suggesting that similarity is an important component in the processing of meaningful material. Chi, Feltovich and Glaser (1981) demonstrated that both novices and experts classified physics problems on the basis of similarity, although the novices focused on surface similarities while the experts focused on similarities in the problem's solution. More recently, Holyoak and Koh (1987) demonstrated that subjects were able to solve problems by recognizing abstract similarities between a new situation and an old situation and applying the solution that was appropriate in the past to the new situation.

Our implementation of Hintzman's model can be seen in Figure 2, which is based on Hintzman's Figure 1 (1986, p. 413). In this model, presentation of a test picture recruits similar stored instances from memory. These instances are activated with a strength A_i proportional to a power of the similarity between the new example and the stored instance. The overall force effectiveness rating for each instance in memory is stored as a feature in the representation of the instance. The inferred values for all unspecified features in the new example, including the force effectiveness rating, is the weighted average of those feature's values in the activated instances.

Insert Figure 2 about here

We should note a minor modification we made to Hintzman's original model. In Hintzman's implementation, features are encoded as all or nothing entities that are absent or present (as 0's and 1's). In our materials, the features are graded (e.g., number of ships). Rather than represent these gradations as a series of binary features (e.g., 0 ships, 1 ship, 2 ships, etc.), we have chosen to represent features in a graded fashion, on a continuum from zero to one rather than discretely as one or the other. Consequently, similarity between a new example and an old instance is computed from the similarity between features. In the Hintzman model, each feature was either present or not and similarity was computed from the number of features in common. In our variant, features differ by their characteristics or "strength" and similarity is computed by summing the closeness in strength of individual features. Appendix A describes the formula which we used to compute the subject's effectiveness ratings predicted by this model.

First enhancement -- extended Hintzman with feature value adjustment

While Hintzman's model has empirical support, situations can be imagined where his model might not produce acceptable predictions. To take a trivial example, suppose that subjects were presented with only a single training picture, which was given a rating of five (on a scale of 1 to 10). Suppose further that a test item is presented which is a stronger example of the category (e.g., one that has twice as many ships, aircraft and submarines). Intuitively, it would seem likely that subjects would produce a rating greater than five for this test picture. This sort of situation cannot be accomodated by Hintzman's model in its present form.

A possible alternative to the Hintzman model which accomodates this situation is that independent estimates of the test exemplar are generated from each similar instance in memory and these independent estimates are averaged. An independent estimate is formed by comparing the new example with the new instance feature by feature. For each feature in the new example that is stronger than that in the stored instance, the effectiveness of the new example is adjusted up; for each weaker feature, it is adjusted down. Note that this adjustment is possible only if feature and overall strength can be scaled and if a meaningful relationship exists between feature characteristic and overall effectiveness. This relationship might exist naturally with "meaningful materials". It will not exist with nonsense cases.

For example, if the test picture has many more ships than the given training picture, then the effectiveness rating for that feature would be scaled upwards. If the test picture had one or two fewer ships than the given training picture, then the effectiveness rating would be scaled slightly downward. The overall force effectiveness rating inferred for the new example is a weighted average of these individual estimates. This model can be seen in Figure 3. In fact, this model might better account for the development of "normative" responses for features (as suggested by Kahneman and Miller, 1986) when the entire range has not been represented during training.

Insert Figure 3 about here

The actual formulae used in computing the predicted scores for this model are also shown in Appendix A. It should be noted that this extension does not add a parameter to the model. The models illustrated by Figures 2 and 3 contain the same number of free parameters.

Second extension -- extended Hintzman model with functional feature substitution

Like the first extension, this one also is motivated by a simple "what if" example. Suppose, for example, that some of the ships in Figure 1 were replaced by some other kind of platform, like a dirigible with missiles. In this case, it is unlikely that people would completely ignore the dirigible in estimating the force's effectiveness. Rather, they would more likely count the dirigible as a ship. In estimating attack effectiveness, they would substitute an appropriate number of ships for the armed dirigible.

Unlike the original model and the first extension, which are evaluated in both experiments, this extension is addressed only in experiment 2.

Summary

This research was designed to examine how well Hintzman's (1986) exemplar-based model can account for subjects' evaluations of meaningful new examples after being trained on similar old examples. In experiment 1, this model is compared with an extension designed to exploit meaningful relationships between the characteristics of features and the overall example. Experiment 2 evaluates a second extension designed to accomodate more complex examples.

Experiment 1: Evaluation of "All-Out Attacks"

Method

Subjects. The subjects were 45 undergraduate students at George Mason University in Fairfax, Virginia. Five of these subjects were unable to accurately predict the overall ratings for 9 of the 12 training pictures within three training trials; their data were not analyzed further. The students received either course credit or payment for their participation in the study.

Materials. The materials for this experiment consisted of two sets of 12 training pictures, two sets of 10 test pictures, a set of feature evaluation sheets, and the Raven Progressive Matrices Test (1958), which was used as a distractor task.

Both sets of training and test pictures illustrated military threats capable of mounting "all-out attacks" with differing degrees of effectiveness. The pictures contain friendly forces (white) and hostile forces (black) which are surrounding the friendly forces. The locations and number of hostile forces vary, but the location and number of friendly forces is constant. Each picture was accompanied by an attack effectiveness rating; that is, a rating of how effective the all-out attack in the picture is, and a feature-based explanation of that rating.

Figure 1 shows a sample picture, its rating, and its explanation for the rating. The first set of materials contained pictures where the number of hostile forces ranged from 7 to 26. The second set was identical to the first set except that each of the training and test pictures in Set 2 contained 50% more platforms (ships, submarines, aircraft) than the corresponding picture in Set 1. The additional platforms were placed in the same quadrant of the picture and close to the original platforms in order to minimize any effect on the perceived number of attack axes. The words used to describe the pictures were identical to those used in Set 1. The ratings were calculated from a formula that was never mentioned to the participants.

The feature evaluation sheets contained the names of the features whose feature effectiveness ratings were used in computing the attack effectiveness ratings for the pictures. For each feature, there were three blanks to be completed: (1) a rating of the extent to which each feature in the accompanying picture is characteristic of an all-out attack (from 1 = characteristic of a very poor all-out attack, to 10 = characteristic of a very good all-out attack); (2) how important that feature would be in estimating the effectiveness of that all-out attack (from 1 = not at all important, to 10 = very important); and (3) how confident subjects were of the ratings they had just assigned for the feature (from 1 = not at all confident, to 10 = very confident).

Procedure. The experiment began with a training session in which the subjects were provided with background material explaining the basic Battle Group scenario with which they would be working. Subjects were then shown six examples of all-out attacks (either from Set 1 or from Set 2). They were told how each picture's attack effectiveness had been rated by an "expert" (it was actually computed by a formula) on a 10-point scale and were given a feature-based explanation of the rating. They were then shown six additional training pictures from the same set and were asked to predict the "expert" rating given for each. After each prediction, the actual rating and feature-based explanation of the rating was provided for the training examples. Subjects cycled through these twelve training pictures until they could accurately predict (within one point) the "expert's" attack effectiveness ratings for 9 of the 12 pictures.

After the training session, subjects were shown ten test pictures (consistent with the set they had studied). For each picture, they were asked to rate how effective the all-out attack pictured was and how confident they were that their rating would match the "expert's" attack effectiveness rating within one point. Each of these judgments was made on a 10-point scale.

When the attack effectiveness ratings had been completed, subjects were asked to work on a series of puzzles, which were designed to serve as a distractor task. After working on these puzzles for twenty minutes, the subjects completed the feature rating sheets for each of the ten test pictures.

After the feature rating sheets had been completed, the subjects were again asked to provide attack effectiveness ratings for the ten test pictures that had been presented earlier. They were also asked to make confidence ratings for each of their judgments.

Results

The subjects were apparently basing their estimates on the training instances presented in this experiment rather than on any specific prior knowledge about how "all-out attacks" work. Two groups of subjects participated in this experiment, the materials and procedures for these two groups were identical, except that the number of ships, submarines, and aircraft in each training picture shown to group 2 subjects was 50% greater than the number in the corresponding picture shown to group 1 subjects. Despite being shown much larger attacking forces, subjects in group 2 did not estimate higher force effectiveness ratings than subjects in group 1. In fact, the overall average rating given by group 1 subjects was 5.98 while that given by group 2 subjects was 5.79.

This result indicates that subjects were calibrating the responses to the instances shown in training. It also supports the Kahneman and Miller (1986) norm theory, in which various qualities in new examples are judged with respect to the norm in previously-seen instances.

Since the subjects were basing the effectiveness estimates on the training examples, we wished to determine if these ratings correlated with the ratings predicted by the original and extended Hintzman models. We also wished to determine which of the models better accounted for the subjects' ratings.

To do this, we computed the subjects'effectiveness ratings predicted by each model, using the formulae shown in Appendix A. These calculations were extensive because the effectiveness ratings predicted by each model depend on the values of its free parameters. To remove any possible bias from the selection of the free parameters, we set the parameters separately for each subject and each model to optimize model performance. The intent of these calculations was to empirically derive the best fit for each subject and compare, for each subject, the best fitting original Hintzman model with the best fitting extended Hintzman model.

The subjects' attack effectiveness ratings for each of the ten test pictures are the principal independent variables. In the experiment, the subjects were asked to provide attack effectiveness ratings on two independent occasions, about 60 minutes apart. These two ratings did not differ significantly ($t(399) = -1.694$, $p > .05$, combining data from the low and high density conditions). Due to this consistency, the attack effectiveness ratings were averaged and this average score was used in subsequent analyses.

Correlations were calculated between each subject's average attack effectiveness rating and the attack effectiveness ratings predicted by each of the two models for that subject; they are shown in Table 1.

Insert Table 1 about here

These correlations were then converted to z-scores using Fishers's r to z transformation (Hays, 1973, p. 661-663) and a z-statistic was calculated. The mean z-score was then converted back to an r value to determine the average correlation for the model.

The resulting z-statistic for the Hintzman model was found to be significant ($z = 9.56$, $p < .01$), with a mean correlation of .92. The correlation for the extended Hintzman model with additional processing was also found to be significant ($z = 10.56$, $p < .01$), with a mean correlation value of .94.

Further analyses were conducted to determine which model was a better fit to the data on a subject-by-subject basis. A stepwise regression was calculated for each subject using the best fits of each model for each subject as independent variables. The first (and only) model entered into the equation is shown in Table 2 for each subject in the low and high density conditions, with their R^2 values. With the exception of subject 17, whose data could not be fit by either model, the fits were fairly good (with R^2 values ranging from .72 - .98 for the first model entered into the equation).

Insert Table 2 about here

Across the low and high density conditions, the extended Hintzman model better predicted subjects' ratings 77% of the time. A sign test for matched pairs showed that the extended Hintzman model was significantly better than original Hintzman model ($z = 2.72$, $p < .05$).

Discussion

This research examined two alternate explanations of how people make judgments about new instances within a domain when that domain contains real-world materials. The first model, based on Hintzman's (1986) multiple-trace memory model, proposes that evaluations of new instances are feature based, and that subjects' estimates are interpolated from previously-seen examples. The second model, also exemplar-based, proposes that subjects use some world understanding to evaluate a new example independently from each of the prior instances and then average these independent estimates. For the materials in this experiment, this independent estimate was accomplished by scaling the attack effectiveness ratings for a given training picture relative to a given test instance based on the direction and amount of dissimilarity between the training and test pictures on each feature.

The goals of this research were to determine a) how well Hintzman's model could explain the process of judging new instances of a category and b) whether the model is improved by permitting it to take into account not only the similarity between a new example and each old instance, but also the significance of the differences between a new example and an old instance. While our data show that the original Hintzman model captured the variance in subjects' evaluations of new instances fairly well, the extended Hintzman model does account for the subjects' judgments somewhat better.

The data thus suggest that examples are not retrieved and used as is. Rather, the subjects may be interpreting each old instance before using it. In the extended Hintzman model, this interpretation is simple -- merely an adjustment of each instance to take into account the degree and direction of feature differences between a new example and the old instances.

Our materials were not designed to provide an advantage to the extended Hintzman model. None of the test pictures lay outside of the range of the training pictures, which included all the attack extremes. That is, none of the attacks in the test pictures were stronger than the strongest training picture attack, nor were any weaker than the weakest training picture attack. Thus, none of the test pictures fit the example described previously, in which subjects provided with a single training picture would likely rate a test item containing twice as many platforms higher than the training picture.

The better results from the extended Hintzman model may thus reflect a fundamental principle of how people use old instances to evaluate a new example: they use world understanding to independently evaluate new instances and then combine these evaluations.

Experiment 2: Evaluations of "Barriers"

The data from the first experiment suggested that a model proposing that subjects generate independent pre-processed estimates from each training example and then average them to generate the rating of the test instances better accounted for judgments about a particular real-world domain than did Hintzman's original (1986) model. This pre-processing enabled the model to reflect certain real world knowledge about how all-out attacks work, e.g. larger sized forces are usually more effective. Our objective in the second experiment was to take this research further and examine another extension of the Hintzman model that enables it to integrate real world knowledge. In this case, the extension is functional feature replacement.

In this study, the situations to be evaluated were "barriers". These barriers were represented in drawings by a row of hostile forces (which varied) which are attempting to block a battle group from progressing towards their destination. Like the all-out attack, these barriers were not intended to be realistic to a military planner. Rather, they were intended to represent simple situations where subjects can apply general knowledge about barriers in evaluating new barrier examples. As in the all-out attack experiment, subjects were initially shown examples of barriers. Each instance included a rating of its example as a barrier and a feature-based explanation of the rating. For example, Figure 4 shows an example of a training picture with its associated rating. Subjects were then shown new instances and asked to evaluated how good these new instances were as examples of barriers.

Insert Figure 4 about here

The goals of this experiment were (a) to replicate experiment 1, testing whether the extension of Hintzman's model will again be the model that better accounts for the way in which people make judgments about new instances, (b) to examine whether an additional extension, "functional substitution", can explain subjects' ratings of new kinds of barriers that were not seen during the training phase, and (c) to determine the range of new barrier types where such an extension can explain subjects' evaluations.

Method

Subjects. The subjects were twenty undergraduate students at George Mason University in Fairfax, Virginia. The students received either course credit or payment for their participation in the study.

Materials. The materials for this experiment consisted of a set of ten training pictures, seventeen test pictures, and the Raven Progressive Matrices Test (1958), which was used as a distractor task.

The training and test pictures for this experiment illustrated situations defined as "barriers" where hostile forces are attempting to prevent a Battle Group from moving forward by blocking the path towards their destination. The set of training pictures and one set of test pictures were constructed from a model of barrier goodness. This model specifies two features relevant for barrier effectiveness assessment, the length of the barrier and the solidity of the weakest part of the barrier and scores the features based on measurable physical attributes. The overall barrier effectiveness was then calculated from the weighted geometric mean of the feature scores obtained from these two features. Since in our model of barrier effectiveness, a barrier was only as strong as its weakest link, the weaker feature was weighted more heavily. For this experiment, we arbitrarily chose to weight the weaker feature by 0.75 and the stronger by 0.25, using these values as exponents in the geometric mean.

For the first part of this experiment, 15 pictures were developed; five pictures were shown during training only, five were shown during test only, and five were shown both as training and test pictures. The overall ratings of the pictures ranged from two to ten for both the training and test pictures.

Seven pictures were developed for the second part of the test. These pictures were modifications of pictures shown in the first part. Five of the pictures were modified by adding either an island or peninsulas to the picture. This procedure created pictures which physically matched one of the original test pictures in terms of number and location of platforms, but functionally matched a second original test picture, in terms of length and solidity of the barrier. Two other new test pictures were created by taking two of the original test pictures and moving the platforms to one side, so that they were no longer centered in front of the battle group. Again, this created pictures which were physically similar to one of the original test pictures, but functionally similar to another original test picture.

Procedure. The experiment began with a training session in which the subjects were provided with background material explaining the basic Battle Group scenario with which they would be working. Subjects were then shown five examples of barriers. They were told how each picture's barrier effectiveness had been rated by our model on a 10-point scale and were given a feature-based explanation of the rating. They were then shown five additional training pictures and were asked to predict the model rating given for each. After each prediction, the actual rating and feature-based explanation of the rating was provided for the training examples. Subjects cycled through these ten training pictures until they could accurately predict (within one point) the model's barrier effectiveness ratings for 8 of the 10 pictures.

After the training session, subjects were shown ten test pictures (five were seen during training and five were new). For each picture, they were asked to rate how effective the barrier pictured was and how confident they were that their rating would match the model's barrier effectiveness rating within one point. Each of these judgments was made on a 10-point scale.

When the barrier effectiveness ratings had been completed, subjects were asked to work on a series of puzzles, which were designed to serve as a distractor task. After working on these puzzles for ten minutes, the subjects were asked to make effectiveness ratings on seven additional pictures.

Results

Replication of Experiment 1. The first question of interest was whether the results of the first experiment would be replicated here, with the enhanced Hintzman model explaining the data better than the original Hintzman model. This question was addressed by the first ten test pictures. As in the first experiment, the first step in the data analysis process was to compute the predictions for the original Hintzman model and the first extension of the model. The general formulae for these models are the same as those used in the first experiment. As in the first experiment, the actual feature sets and similarity exponent used by the subjects were not known; therefore, we examined a number of different possible feature sets and similarity exponents. For each subject and each model, we selected the set that provided the best fit to the data. Again, our intent was to avoid inadvertantly biasing our results by picking free parameters more favorable to one model than the other. Therefore, we compared for each subject the best fitting Hintzman model with the best fitting extended Hintzman model.

Correlations were calculated between each subject's average barrier effectiveness rating and the barrier effectiveness ratings predicted by each of the models for that subject for the first ten pictures (i.e., the ones not involving new information); they are shown in Table 3.

Insert Table 3 about here

These correlations were then converted to _z_-scores using Fisher's _r_ to _z_ transformation (Hays, 1973, p. 661-663) and a _z_-statistic was calculated. The resulting _z_-statistics were found to be significant ($z = 6.68, 7.22$, $p < .01$ for the Hintzman and extended Hintzman models, respectively). The mean _z_-scores were then converted back to _r_ values to determine the average correlation for each model (Hintzman = .92 and extended Hintzman = .94).

Further analyses were conducted to determine which model was a better fit to the data on a subject-by-subject basis. A stepwise regression was calculated for each subject using the best fits of each model for each subject as independent variables. The first model entered into the equation is shown in Table 4 for each subject, with their R^2 values. The fits were again fairly good (with the exception of one subject whose data could not be fit by either model). The R^2 values (excluding the subject whose R^2 value was .45) ranged from .69 - .96 for the first model entered into the equation.

Insert Table 4 about here

Again, the extended Hintzman model better accounted for the data. Eighty-four percent of the time, this model was entered into the regression equation first. A sign test for matched pairs showed that the extended Hintzman model was significantly better than the original Hintzman model ($z = 2.46$ $p < .01$).

Extension to Functional Feature Substitution. The second question of interest was whether the Hintzman model could be further extended to functional substitution of features. With this extension, people would substitute features of a new exemplar with functionally equivalent features in previously-seen exemplars. Of course, such functional substitution requires that people be able to evaluate the functionality of features.

During the training for this experiment, subjects only saw barriers that were a centered row of ships or submarines. They never saw barriers that were off-centered or that contained islands or peninsulae. During the test, pictures with islands or peninsulae, or that were off-centered, were shown in order to evaluate the functional substitution extension of the Hintzman model. This extension proposed that people would use general knowledge about islands and peninsulae (ships cannot pass over land) and off-centeredness (easier to go around) in making their feature assessments and in evaluating the quality of the barriers.

To compute the feature values needed by the formulae in Appendix A for the functional substitution extension to the model, we needed objective methods for converting the new types of examples to functionally equivalent examples of the type of barriers seen in training. Islands and peninsulae were substituted by a row of ships, whose length was equal to the length of the island or peninsulae. Functional gap size was computed by calculating the resulting physical gap size had the islands been a row of ships. For off-centered barriers, we computed a functional length from how hard the barrier would be to go around. This length was computed by projecting the intersection of the battle group's path with the barrier, finding the distance between that intersection point and the nearest end, and doubling this distance.

Subjects' representations for barriers accomodated islands and unusual barrier placement. Subjects had no trouble giving ratings for these new types of barriers. In all cases, adding islands or peninsulae or moving the barrier to an off-centered location had a substantial effect on the subjects' effectiveness ratings.

Taking all the pictures as a group, the Hintzman model with functional replacement did not explain the subjects' ratings. Correlations were calculated between each subject's average barrier effectiveness rating and the barrier effectiveness ratings predicted by each of the models for that subject for the second seven pictures (i.e., the ones involving new information). Table 5 shows these correlations for the original Hintzman model and the first extension, both with and without functional substitution.

Insert Table 5 about here

These correlations were then converted to *z*-scores using Fisher's *r* to *z* transformation (Hays, 1973, p. 661-663) and a *z*-statistic was calculated. The resulting *z*-statistics for the original Hintzman model and the first extension to the model were found to be non-significant both without functional feature substitution ($z = 1.45$ and 1.73, $p > .05$ for the Hintzman and extended Hintzman models, respectively) and also with this substitution ($z = 1.58$ and 1.71, $p > .05$ for the Hintzman and extended Hintzman models, respectively). The mean *z*-scores were then converted back to *r* values to determine the average correlation for each model (Hintzman = .34 and extended Hintzman = .40 without functional substitution and Hintzman = .37 and extended Hintzman = .39 with functional substitution).

These low overall correlations do not necessarily imply that functional feature substitution does not take place. It may only mean that the functional substitution rule that we selected is sometimes not appropriate. Table 6 provides a schematic of each test picture and summarizes the different barrier effectiveness ratings: the original formula rating used to construct the pictures, the predicted rating from the first extension to Hintzman's model, and the predicted rating from the extension to Hintzman's model using functional substitution on a picture-by-picture basis, averaged across subjects.

Insert Table 6 about here

An examination of the results for the different pictures suggests that subjects may indeed have been performing a form of functional substitution, but in some cases, were estimating functional equivalents using more sophisticated and complex methods than our simple substitution rules. Functional substitution worked for three of the four island cases (pictures 11, 12, and 14) and for the peninsula (picture 15). It did not work, however, for picture 13. This barrier had two islands in the center which, in retrospect, we presume subjects assumed could provide a safe passage for the ships; however, its functional equivalent had been set as a very long barrier. Functional substitution also did not work for the off-centered barriers (pictures 16 and 17). Subjects did not seem confident of the battle group's ability to go around, as they rated these barriers higher in quality than their presumed functional equivalents. Perhaps they thought that the barriers would reposition themselves as the battle group progressed.

Discussion

This study was designed partly to replicate experiment 1 and partly to test a second extension to the Hintzman model. The results for pictures 1-10, which resembled the training pictures physically, were very similar to those in experiment 1. Both the original and extended Hintzman models accounted for subjects' judgments about new instances, but the extended model did somewhat better.

When new information was added to the pictures (e.g., in the form of islands), neither of the models evaluated in experiment 1 provided a good fit to the data. When the Hintzman model is extended further using functional substitution, it was able to account for subjects' ratings in four of seven cases. This result supports the idea that subjects integrate relevant world knowledge into their judgments about new exemplars. In rating pictures containing features never seen during training, subjects were able to translate the newly-presented physical features into equivalent functional features (e.g., extending the barrier length when a peninsula, rather than ships, was added to the picture). Interestingly, this extension did not work for three of the test cases. It is possible that these cases could also work within the basic model concept, but require more sophisticated methods for functional substitution.

General Discussion

The results from the two experiments, taken together, suggest that Hintzman's (1986) model can provide a good foundation for understanding how people judge instances from a category using real world materials. The results also suggest that while this model in its original form has limitations, it can be readily extended to accomodate more complex situations. These extensions concern ways in which general world knowledge (more ships increase attack strength, ships don't sail through islands) can augment previously seen instances in evaluating new examples.

The results from the second experiment further suggest that the simple mechanisms outlined in this study for incorporating world knowledge into the model calculations (that is, substituting new features with familiar functional equivalents) do not capture some important processes used by subjects to evaluate new instances. While the model predictions were consistent with the ratings for four of the new instances, they were inconsistent for the other three pictures.

These results have implications for the nature and acquisition of expertise. As extended, the Hintzman model describes the cognitive basis for expertise. People become more expert in a domain as they acquire more examples in that domain and as they identify more useful features for representing each example in the domain. This research suggests that having more examples improves performance because a new case is more likely to match closely one of the old examples. The set of features used also affects expertise because the features affect the quality of the match assessment and because they are used to evaluate the significance of differences between a new example and previously experienced examples. Expertise depends not only on acquiring more examples, but also on learning what features should be used for representing each example.

Should this model of expertise be correct, it can be used to help train experts. Expertise is developed by practicing examples. It may be developed more rapidly if these examples are selected to emphasize those features that matter most in evaluating new cases and if the training makes clear the relationship between these features and other important qualities in the example.

References

Alba, J. W. and Hasher, L. (1983) Is memory schematic? Psychological Bulletin, 93, 203-231.

Brewer, W. F. and Nakamura, G. V. (1984) The nature and function of schemas. In R. S. Wyer and J. K. Scull (Eds.), Handbook of Social Cognition (pp. 119-160). Hillsdale, N.J.: Lawrence Erlbaum.

Brooks, L. (1987) Decentralized control of categorization: the role of prior processing episodes. In U. Neisser (Ed.) Concepts and conceptual development: Ecological and intellectual factors in categorization (pp. 141-174). New York: Cambridge University Press.

Chi, M. T. H., Feltovich, P. J., and Glaser, R. (1981) Categorization and representation of physics problems by experts and novices, Cognitive Science, 5, 121-152.

Hays, W. L. (1973) Statistics for the social sciences. New York: Holt, Rinehart and Winston, Inc.

Hintzman, D. L. (1986) "Schema abstraction" in a multiple-trace memory model, Psychological Bulletin, 93, 411-428.

Holyoak, K. J. and Koh, K. (1987) Surface and structural similarity in analogical transfer, Memory and Cognition, 15, 332-340.

Kahneman, D. and Miller, D. T. (1986) Norm theory: Comparing reality to its alternatives, Psychological Review, 93, 136-153.

Raven, J. E. (1958) Standard progressive matrices, sets A, B, C, D, and E. London: H. K. Lewis and Company, Ltd.

Whittlesea, B. W. A. (1987) Preservation of specific experiences in the representation of general knowledge, Journal of Experimental Psychology: Learning, Memory and Cognition, 13, 3-17.

Acknowledgements

The authors would like to thank Steve Masterson for his help in data collection. This research was sponsored by the Office of Naval Research, Engineering Psychology Program (Contract N00014-84-C-0484). The views expressed in this paper are not necessarily those of the Office of Naval Research or the Department of Defense.

Appendix A

Formulae for Hintzman Model

Three formulae are used to compute the model predicted ratings for each picture. The first formula defines the similarity of a feature (i) in a test picture (k) relative to the same feature in a training picture (j). Similarity is computed as:

$$FEAT\ SIM_{ijk} = 1 - ABS\left(\frac{featvaltrain_{ij} - featvaltest_{ik}}{(\text{range of } featval_i)}\right) \quad [1]$$

That is, the difference between a given feature (i)'s value (e.g., the number of ships in an "all-out attack") in a particular training picture (j) and in a particular test picture (k) was divided by the range of values that the feature (i) could assume. This difference could be raised to a power. Since sensitivity tests show that the value of this parameter did not matter, it was set to unity (1). The absolute value of this difference is then subtracted from 1 to provide a similarity rating between 0 and 1. The similarity rating for two identical features in two different pictures is unity. The rating is zero only when the values of the features being compared are at different extremes of the value scale.

The second formula defines the overall similarity of a particular test picture (k) to a particular training picture (j) based on all features deemed relevant. It is a weighted average of the FEAT SIM (ijk) over all features. Because feature weight is not measured, the PIC SIM was calculated using different weights (as outlined in Table 7) which represent the importance of the feature (i) to the task.

$$PIC\ SIM_{jk} = \frac{\text{Sum over features } (featsim_{ijk}) * (featwt_i)}{\text{Sum over features } (featwt_i)} \quad [2]$$

Finally, the predicted rating for each test picture (k) is calculated by using a weighted average of the briefed ratings of the training pictures.

$$\text{PRED EFF SCORE}_k = \frac{\text{Sum over train pic } ((\text{PicSim}_{jk})^P * \text{brfd eff}_j)}{\text{Sum over training pic } (\text{PicSim}_{jk})^P} \quad [3]$$

The picture weight is the picture similarity value calculated in equation [2] raised to a power. Here, P is a free parameter relating physical similarity to subjective similarity. The PRED EFF SCORE was computed iteratively, with P being varied between 1 and 20.

It should be noted that formulae [1] and [2] require values for the relevant features and for the power P that weights similarity. In the case of our materials, the formula used to compute the briefed effectiveness ratings used certain assumed features. However, it was not clear that subjects would choose the same features or weight them in the same way as we had. Therefore, we could not be certain that these features should be the ones used in equations [1] to [3]. Nor did we have any reason to assume that all of the subjects would make their judgments based on the same features or give them the same weights. In fact, the subjects were not consistent in their ratings of the importance of features across pictures. Consequently, the features used in equations [1] to [3] might be differnt for different subjects. In order to ensure that the feature sets and power P chosen for equations [1] to [3] did not inadvertantly favor one model over the other, we adopted a procedure that compares the best possible version of the two models. Thus, we examined several different feature sets and powers P and selected, for each subject and each model, the feature set and power that maximized the variance accounted for. While not all subjects had the best fit for both models with the same feature set, most subjects were fairly consistent.

In choosing the sets for the all-out attack evaluation, the six features used in the original design of the training and test pictures were considered as well as two additional features (total number of platforms and overall surroundedness). These potential eight features are listed in Table 7. The features were given weights of 0, 0.5, or 1 as shown in the table. The ten feature sets shown in the table were selected as the most promising from more than 20 that were tested with the mean data.

Insert Table 7 about here

Formulas for Extended Hintzman Model with Additional Processing

This class of models used equations [1] and [2] as described for the previous model. In addition, a formula was needed to allow for the adjustment of the picture ratings in calculating the effectiveness of a test picture (k) based on a given training picture (j). The formula is:

$$\text{PIC PRED EFF SCORE}_{jk} = \text{Briefed Effectiveness}_j + \text{ADJUSTMENT}_{jk} \quad [5]$$

where the adjustment is:

$$\frac{\text{Sum over features }(\text{featwt}_i)(9/\text{featrange}_i)(\text{ftrtest}_{ik} - \text{ftrtrn}_{ij})}{\text{sum}(\text{featwt}_i).} \quad [6]$$

Here, an estimate of picture k's rating is calculated solely from picture j's rating and an adjustment. This adjustment takes into account knowledge of the range of possible feature values and the direction of the effectiveness change from an increase or decrease in feature value. In the formula, the difference in the feature value between the test picture (k) and a given training picture (j) was first multiplied by 9 and divided by the feature range in order to scale the differences between 1 and 10.

The Pic Pred Eff Score was computed as a weighted average of this number. The weights used were the feature weights. The predicted rating for a given test item was then calculated using the general formula:

$$\text{PRED EFF SCORE}_k = \frac{\text{Sum over train pics}((\text{PicSim}_{jk})^P * \text{PIC PRED EFF SCORE}_{jk})}{\text{Sum over training pic }(\text{PicSim}_{jk}).} \quad [7]$$

This predicted effectiveness score for test picture k is a weighted average of the estimates (PIC PRED EFF SCORE) computed from each training example. As in formula [3], the weight is the picture similarity score is raised to a power between 1 and 20.

Note that the two models are closely related. They both have the same free parameters: choice of feature set and power P. The second model differs only because "PIC PRED EFF SCORE" in equation [7] is substituted for "briefed eff" in equation [3].

Table 1. [illegible]
[illegible]dels.

	LOW DENSITY CONDITION		HIGH DENSITY CONDITION	
Subject	Histogram	Extended Histogram	Histogram	Extended Histogram
1	.97	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	.95	[illegible]	[illegible]	[illegible]
4	[illegible]	[illegible]	[illegible]	[illegible]
5	[illegible]	[illegible]	[illegible]	[illegible]
6	[illegible]	[illegible]	[illegible]	[illegible]
7	[illegible]	[illegible]	[illegible]	.94
8	.95	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]	[illegible]
10	[illegible]	[illegible]	[illegible]	[illegible]
11	[illegible]	[illegible]	[illegible]	.95
12	[illegible]	[illegible]	[illegible]	.96
13	[illegible]	.95	[illegible]	[illegible]
14	[illegible]	[illegible]	[illegible]	[illegible]
15	[illegible]	[illegible]	[illegible]	[illegible]
16	[illegible]	[illegible]	[illegible]	[illegible]
17	[illegible]	[illegible]	[illegible]	.95
18	[illegible]	[illegible]	[illegible]	[illegible]
19	.97	.95	[illegible]	[illegible]
20	[illegible]	[illegible]	[illegible]	.94

Table 2. Model accounting for significant (p < .05) variance in each subject's ratings.

Subject	LOW DENSITY CONDITION Model	R^2	HIGH DENSITY CONDITION Model	R^2
1	E	.74	E	.94
2	E	.95	E	.83
3	E	.75	E	.75
4	E	.88	E	.89
5	H	.79	E	.86
6	H	.91	E	.98
7	E	.79	E	.89
8	E	.96	E	.81
9	E	.87	E	.95
10	H	.83	H	.83
11	E	.96	E	.90
12	E	.96	E	.92
13	E	.91	E	.86
14	E	.90	H	.84
15	E	.96	E	.87
16	H	.90	E	.87
17	-	---	H	.72
18	H	.82	H	.81
19	E	.91	E	.93
20	E	.91	E	.88

Table 3. Correlations between actual ratings and two models for barrier pictures 1-10

Subject	Hintzman	Extended Hintzman
1	.98	.98
2	.89	.90
3	.95	.97
4	.87	.88
5	.92	.92
6	.84	.89
7	.92	.96
8	.98	.98
9	.93	.95
10	.35	.37
11	.96	.96
12	.95	.98
13	.94	.95
14	.97	.98
15	.45	.43
16	.82	.83
17	.97	.97
18	.93	.96
19	.95	.96
20	.85	.89

Table 4. Comparison of Hintzman and Extended Hintzman Models for barrier pictures 1-10

Subject	First Model in Equation	R^2
1	H	.96
2	E	.81
3	E	.93
4	E	.78
5	H	.85
6	E	.79
7	E	.91
8	H	.96
9	E	.90
10	E	.76
11	E	.92
12	E	.96
13	E	.91
14	E	.96
15	-	---
16	E	.69
17	E	.94
18	E	.92
19	E	.95
20	E	.79

Table 5. Correlations for each model using physical and functional features for pictures 11-17.

	WITHOUT FUNCTIONAL SUBSTITUTION OF FEATURES		WITH FUNCTIONAL SUBSTITUTION OF FEATURES	
Subject	Hintzman	Extended Hintzman	Hintzman	Extended Hintzman
1	.47	.50	.38	.39
2	-.28	-.26	.14	.12
3	.12	.22	.65	.66
4	.75	.84	.12	.21
5	.34	.47	.61	.66
6	.52	.67	.35	.43
7	.24	.39	.16	.25
8	.04	.16	.53	.54
9	.31	.42	-.10	-.00
10	.37	.29	-.41	-.35
11	.28	.38	.36	.43
12	.50	.56	.22	.30
13	-.00	.21	.60	.70
14	.35	.49	.32	.39
15	.08	-.05	-.34	-.31
16	-.02	.18	.58	.61
17	.72	.76	-.11	-.03
18	.38	.52	.42	.50
19	.28	.44	.76	.82
20	-.05	.03	.75	.73

Table 6. Picture Schematics and Picture Ratings for Barrier Pictures 11-17.

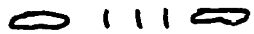
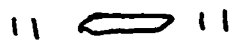
Picture Number	Picture Schematic	Actual Rating	Extended Hintzman Model Predicted Rating	Extended Hintzman with Functional Substitution Model Predicted Rating
11		6.95	4.97	7.05
12		6.70	4.89	6.87
13		3.90	2.19	7.62
14		8.40	5.32	8.58
15		8.35	2.21	9.69
16		6.50	2.61	2.61
17		4.80	2.12	2.12

Table 7. The Weights of Features in Feature Sets

Feature	Feature Set 1	2	3	4	5	6	7	8	9	10
Ships:										
Number	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Directions[1]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Aircraft										
Number	1.0	1.0	1.0	1.0	1.0	0.5	0.5	0.5	0.5	0.5
Directions[1]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Submarines:										
Number	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Quadrants[2]	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0
All Platforms										
Total Number	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0
Quadrants[2]	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.5	0.5

[1] The number of directions from which the enemy is approaching
[2] The number of non-empty quadrants

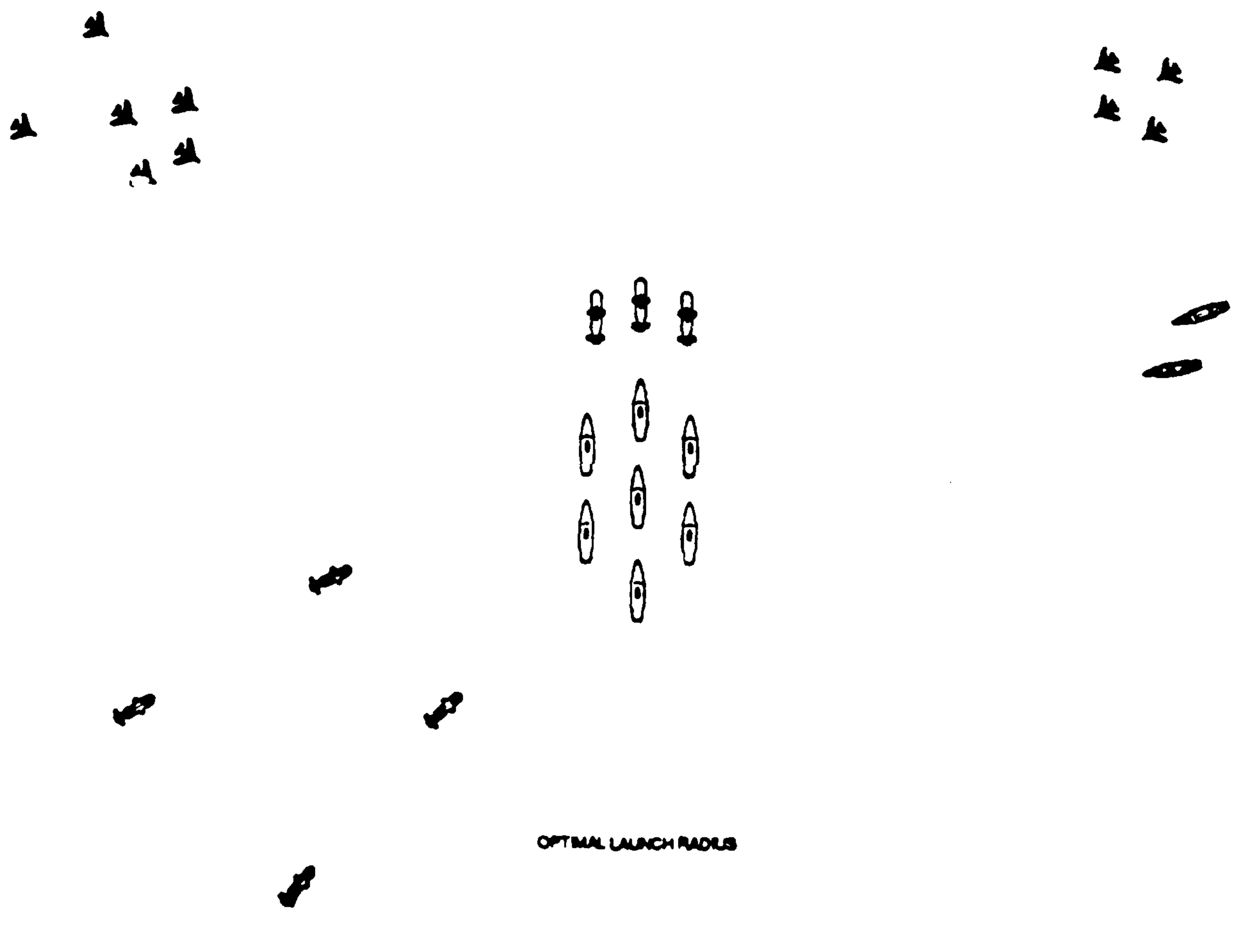
Figure Captions

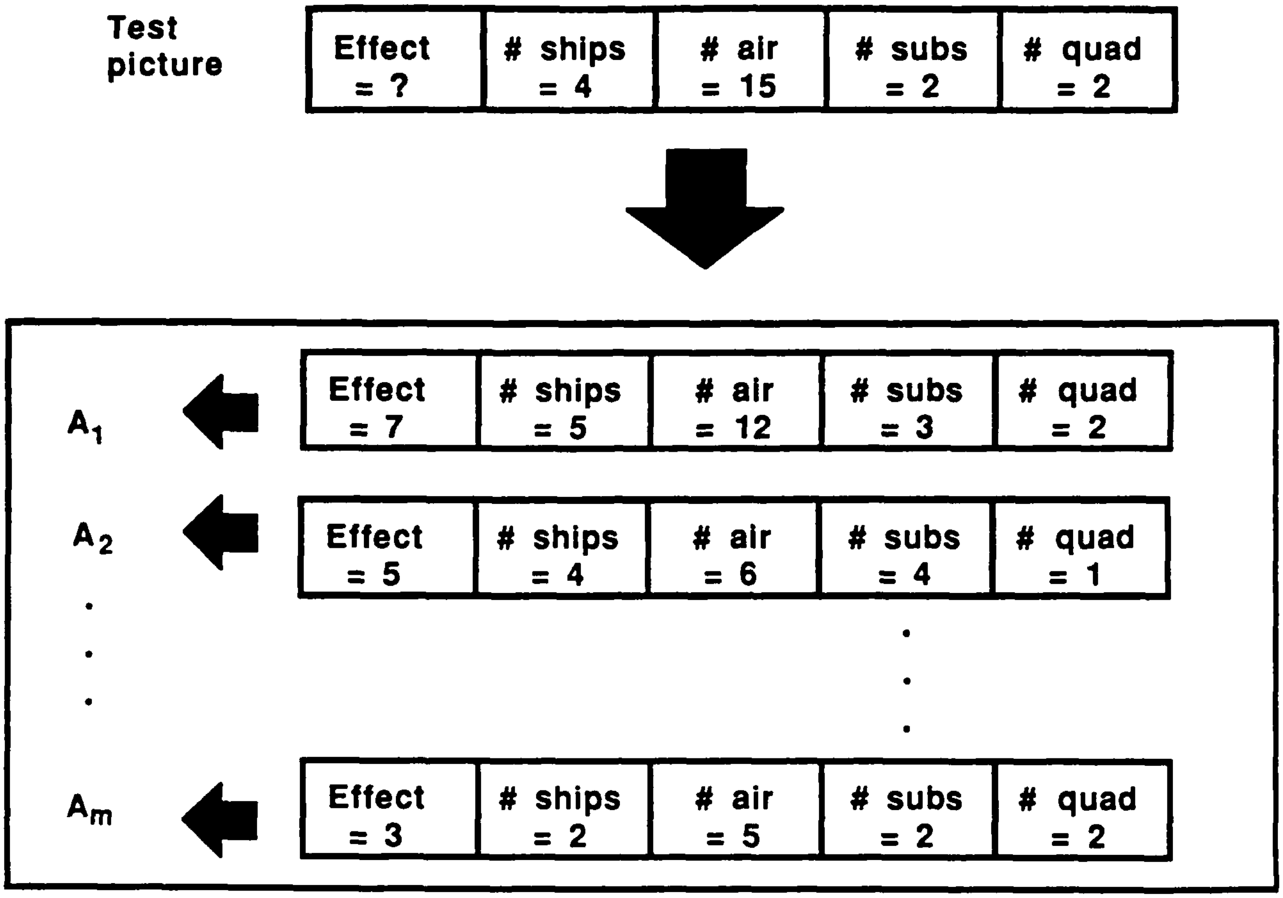
Figure 1. An example of a training picture for all-out attacks. Subjects were told that "Attack effectiveness is 4. The air threat is severe, but the ship and sub threats are weak. There are too few ships, and the submarines are concentrated in only a single quadrant. Note: Attack effectiveness was rated on a scale of one to ten. A score of ten indicated that an all-out attack by these forces would be very effective. A score of one indicated that the all-out attack would be very ineffective.

Figure 2. Summary of Hintzman model. A new instance activates old examples with a strength A_i proportional to the similarity between the new instance and the trace. The inferred value of each unspecified feature, including attack effectiveness, is the weighted average of that feature value in the traces.

Figure 3. Summary of Hintzman model with additional processing. Like the original Hintzman model, a new instance activates old examples with a strength A_i. Each activated trace provides an independent estimate of the effectiveness of the new example based on the effectiveness of the activated trace and the significance of differences between the new example and the activated trace. The inferred effectiveness of the new example is the weighted average of these independent estimates.

Figure 4. An example of a training/test picture for barriers. Subjects were told that "Barrier effectiveness is 10. The barrier is both long and solid. The ships at the two ends are sufficiently far apart to tmake the barrier difficult to go around. The platforms are close enough together throughout its entire length to make passage through the barrier very difficult." Note: Barrier effectiveness was rated on a scale of one to ten. A score of ten indicated that a barrier would be very effective. A score of one indicated that the barrier would be very ineffective.


OPTIMAL LAUNCH RADIUS
MAXIMUM LAUNCH RADIUS



$$\text{Attack Effectiveness} = \sum A_i \,\text{Effect}_i$$

Figure 2. Summary of Hintzman model. A new instance activates old examples with a strength A_i proportional to the similarity between the new instance and the trace. The inferred value of each unspecified feature, including attack effectiveness, is the weighted average of that feature value in the traces

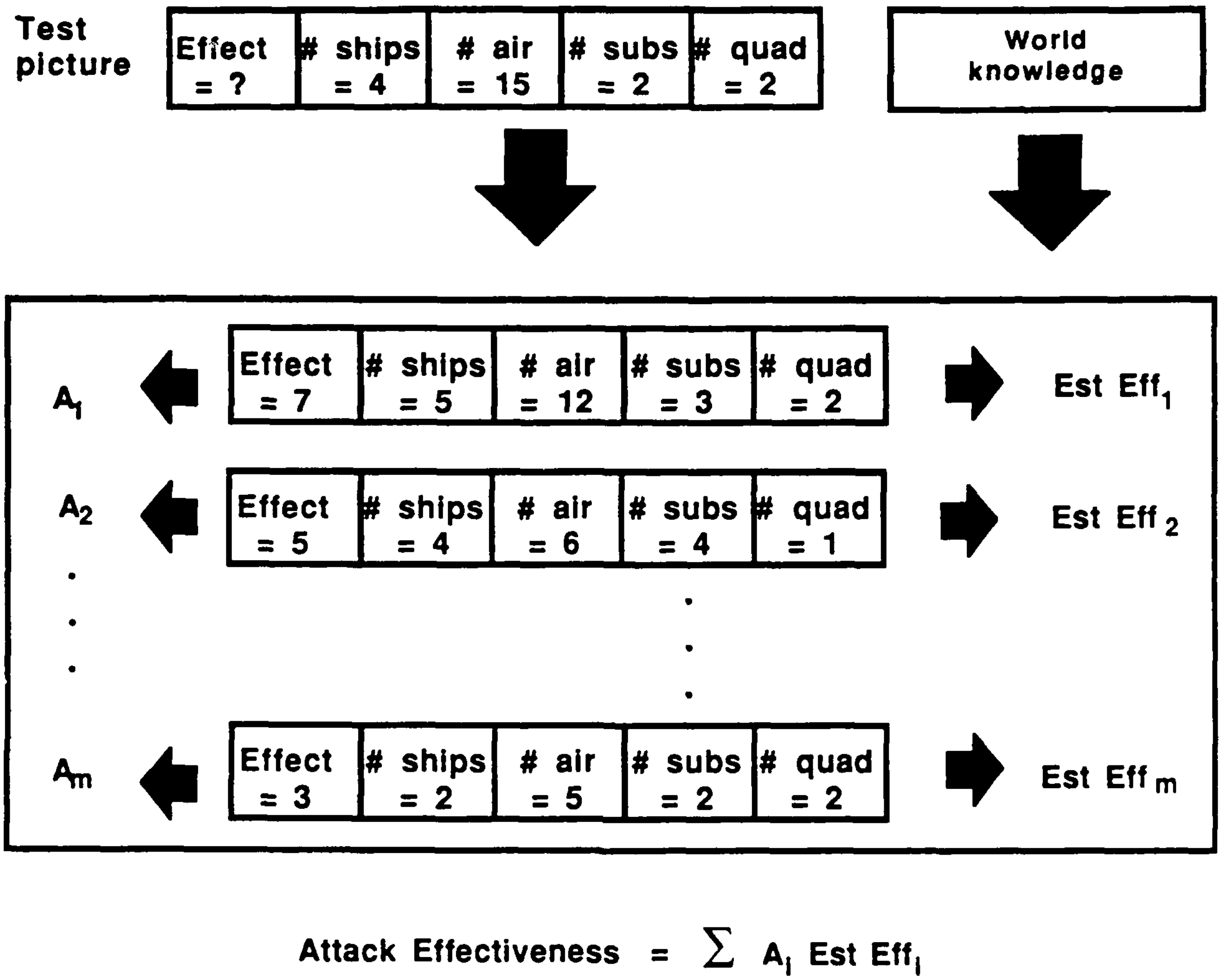
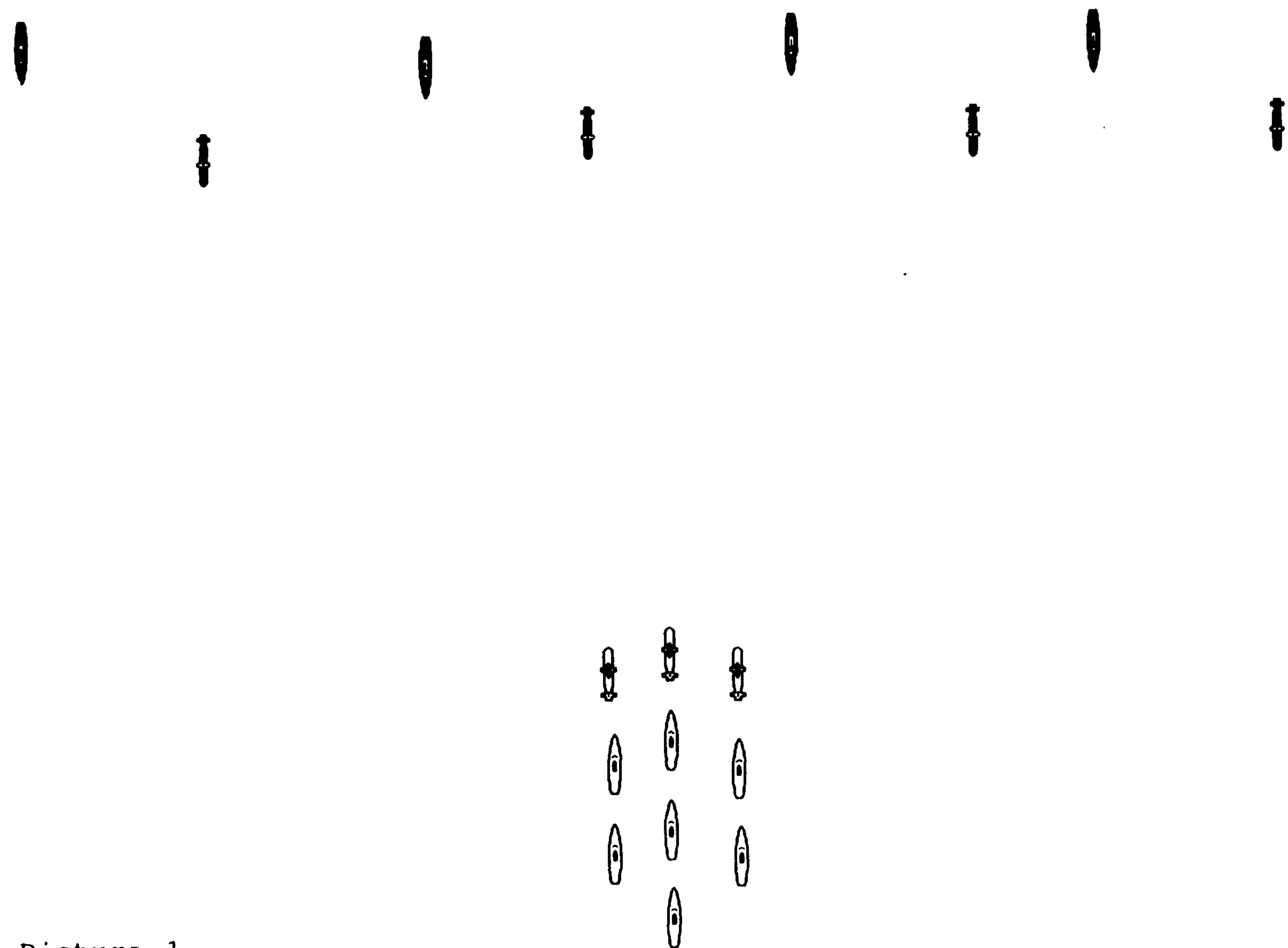


Figure 3. Summary of extended Hintzman model. Like the original Hintzman model, a new instance activates old examples with a strength A_i. Each activated trace provides an independent estimate of the effectiveness of the new example based on the effectiveness of the activated trace and the significance of differences between the new example and the activated trace. The inferred effectiveness of the new example is the weighted average of these independent estimates.



Picture 1

ATTACHMENT 3

THE INTERACTION OF RECOGNITION AND OUTCOME CALCULATION IN DECISION MAKING

Engineering Research Associates
1595 Springhill Road
Vienna, Virginia 22180

OCTOBER 15, 1989

Sponsored by

COGNITIVE AND NEURAL
SCIENCES DIVISION
Office of Naval Research
Contract N00014-84-0484

THE INTERACTION OF RECOGNITION AND OUTCOME CALCULATION IN DECISION MAKING

RECOGNITION AND OUTCOME CALCULATION IN DECISION MAKING

Recognition and outcome calculation are the basis for two contrasting methods of making decisions. When making decisions based on outcome calculation people first identify several promising decision alternatives. They then project the consequences of each alternative, evaluate the desirability of these outcomes, and pick that alternative with the most desirable consequences. Outcome calculation decision making can be characterized as "analytical" because the decision maker explicitly considers factors important in predicting the consequences of different alternative decisions, perhaps employing formal decision-analytic methods or computer simulations. This decision making method has been studied extensively in the past, and has provided a theoretical base for many decision aids.

In recognition-primed decision making, people recognize that a new situation resembles previously encountered situations sufficiently well so that actions that worked in these previously encountered situations are likely to work in the new situation. People who use this method might summarize their decision rationale by saying "I've been in this type of situation before and at that time I took this action. Since it worked then, I will take a similar action this time." Recognition-primed decision making can be characterized as "intuitive." In its most extreme form the decision maker may not be aware of estimating the consequences of different decision alternatives and may not even know what factors influenced his choice. Rather the decision maker just "senses" the right decision.

In many practical decisions, outcome calculation and recognition interact. Over the past several years Engineering Research associates has investigated this interaction, searching for answers to such questions as:

- If an individual is trained to make decisions based on a complex outcome calculation and rule, will he continue to calculate outcomes forever? Or will he eventually, with experience, evolve toward a recognition-primed approach?

- If recognition-primed methods begin to replace explicit outcome calculation, then what kind of knowledge in memory will arise to support recognition-primed decision making? How will this knowledge relate to the specific instances seen in training and how will it relate to outcome-oriented procedural knowledge? Will it be the beginning of production rules, with specific situation indicator / counterindicators that point to an alternative?

- Will people replace outcome calculation with situation recognition when time or resource constraints make projecting outcomes impractical? If this happens, will the recognition-primed decision making completely displace the outcome calculation decision making, or will pieces of recognition-primed decision making become integrated with pieces of outcome oriented decision making?

Two experiments performed by ERA suggest interesting answers to these questions. The results of these experiments indicate that recognition-primed decision making will displace outcome calculation as people become experienced in a decision making domain and that pieces of recognition-primed decision making can be embedded within a larger overall outcome calculation decision process. The results also suggest that outcome calculation and recognition can interact in subtle ways, with outcome calculation influencing the basic cognitive processes used in recognition.

EXPERIMENT 1: HOW TIME PRESSURE AFFECTS DECISION MAKING BASED ON OUTCOME CALCULATION

In this first experiment subjects were trained to evaluate possible decision alternatives by projecting the outcomes of these alternatives. Subjects were given measuring rulers and a mathematical rule to help them compute these outcomes. They were not explicitly taught to recognize situations. During testing subjects lost their rulers and were forced to make decisions much faster than they could if they followed the formal procedures for computing outcomes.

We wished to examine how subjects would react when they could no longer apply the formal procedures taught to them during training. We considered four general possibilities:

1. Subjects would give up, and just guess at the answer

2. Subjects would adapt a wholistic recognition-primed decision making method.

3. Subjects would approximate the formal rule with a very simple one that did not require them to estimate the number of hits from individual ships.

4. Subjects would approximate the formal rule with a quick "eyeball and count" process that included estimating hits from individual ships.

Description of the task

Each subject played the role of a Battle Group commander who encounters a hostile barrier. The subjects were shown a picture of the barrier (Figure 1), which shows the position of hostile ships and the two permitted paths through the barrier. The Battle Group is located at the "X." Each subject was asked to decide whether to traverse the barrier along the straight path, traverse it along the curved path, or stay where he is. He was told to select the path where he receives the fewest hits, unless that number is more than four. In that case, he should stay where he is.

Subjects were taught how to calculate the number of hits likely to be received by a ship as it travels along each path. The computation rules are straight forward. A hostile ship can strike anywhere along the path. As the subject's Battle Group moves along the path, the hostile ships move toward the path. When the hostile ship is at the closest point of approach with respcct to the subject's ship (both are on a line drawn perpendicular to the path being traversed) then the subject calculates the number of hits by measuring the distance between the hostile ship and the path. Ships within the "two hit" distance of the line score two hits; those within the "one hit" distance score one hit. Subjects had special measuring rulers that specified how far ships can move in an hour and that showed the one hit and two hit ranges of the hostile ships.

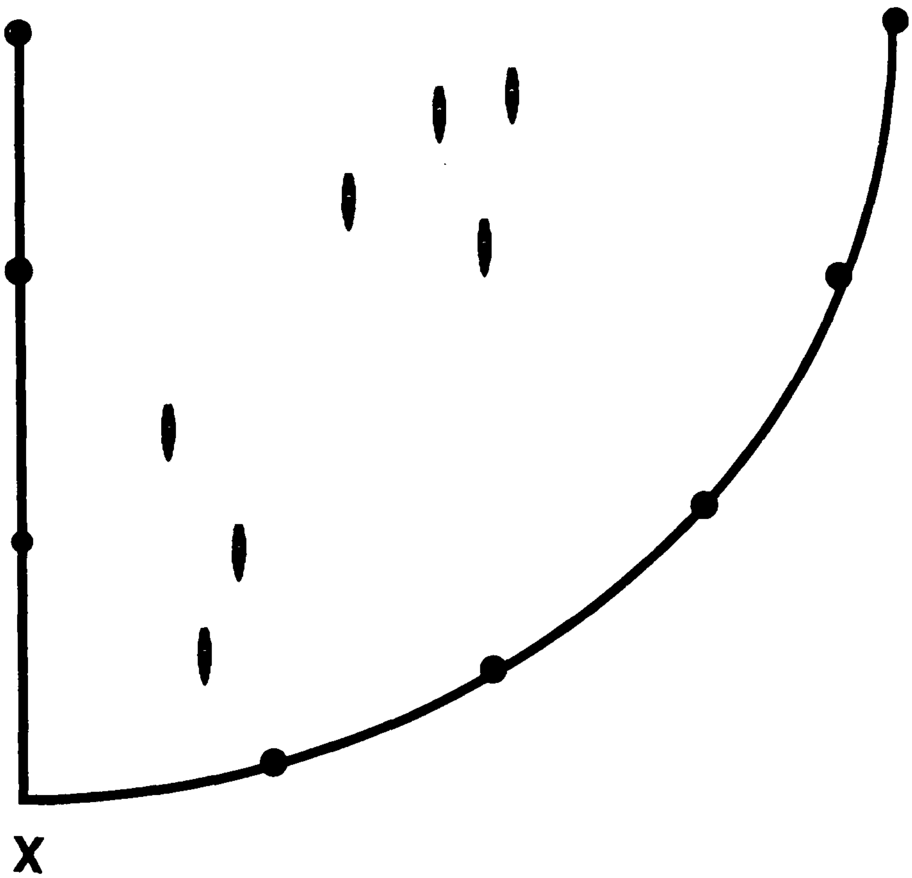


Figure 1. Sample picture of barrier situation (experiment 1).

During the initial training subjects used their rulers to measure the ship movements and to determine ship hits for each path. After the initial training, subjects were provided with a second set of pictures and were asked to rank the three options (traverse along straight path, traverse along the curve path, or stay) without using their measuring tool. They then checked their answers using an answer sheet that gave the correct answers and showed the ship movements.

The test portion of the experiment had seven parts. In part I subjects ranked the "straight," "curved," and "stay" options for each picture in the test set. Each of these pictures were projected for 10 seconds. There was no reminder in the projected picture of the ship movement distances or of the one and two hit ranges. In parts II and III subjects rated each path on a scale of 1 to 10 with respect to "how good the path was at blocking the Battle Group" (part II) and on how well the picture fit the statement "there are many ships near the (_____) path" (part III). In part IV the single ship pictures were projected for 8 seconds each. Subjects entered on the answer sheet their confidence that the shown ship could score '0', '1', or '2' hits. In part V subjects were provided with paper showing only the curved and straight paths, and were asked to draw contours separating the areas from which ships could score zero, one, or two hits. Part VI repeated part I (ranking the options) and part VII repeated part IV (recording the ship hits for the single ships).

Results

Subjects did not just guess at the correct alternative

Subjects ranked the options much better than they would have had they just guessed. Subjects picked the correct option about 70% of the time, a much higher fraction than the 33% expected had they been just guessing. Since the subjects did not have the measuring tools, and would not have had time to measure in any case, they could not have estimated the best option using the formal measurement methods taught in training.

Subjects did not make a wholistic judgment which excluded conscious estimates of the hits from individual ships.

Were they making a "wholistic" decision subjects would feel as if they just "sensed" the right choice from the overall look of the picture, and would not be aware of consciously attending to the picture's detailed components. Presumably this mode of decision making would develop as subjects become so experienced at this task that they would remember the correct choices associated with different types of barriers. When they saw a barrier similar to one that they had previously seen, they would simply select the option associated with that previously seen barrier. There would be no need to compute an estimated outcome. When we first performed this experiment, we hoped that the subjects would make their decisions this way, for this would document a clear case of recognition-primed decision making.

We used subjects' qualitative path assessments to determine whether they based their selections on the overall look of the picture. In parts II and III of the experiment, subjects rated the paths according to "how good the path was at blocking the Battle Group" and how consistent the paths in each picture are with the statement "many ships are near the (straight, curved) path." If subjects based their decisions on the overall look of the barrier, then the ratings given in part II and III should predict the decisions they made in parts I and VI

For each subject, the order of the rankings for the straight and curve options was compared with the order predicted by the qualitative path rankings of parts II and III and with the order predicted by the subject's ship hit estimates given in parts IV and VII. In about 500 cases, the ship hits estimates predicted decisions that differed from the qualitative judgments attained in parts II and III. In these cases, the ship hits predicted the subjects' actual choice 69% of the time, while the qualitative wholistic judgments accounted for it only 31% of the time.

Although these data indicate that our subjects were not basing their decisions on the overall look of the picture, it remains possible, of course, that with enough experience they would start to do so. The conditions in this experiment did not encourage this mode of decision making because subjects did not receive much training in this task, because during training there was no reinforcement of patterns, and because the difference between options was small.

Subjects did not use very simple rules to approximate the outcome calculation.

By simple rules we mean rules that did not require the subjects to estimate the number of hits from each hostile ship. Subjects did not seem to be using rules that were this simple. To evaluate this possibility we calculated the number of correct choices that subjects would have made had they followed a number of different simple rules. For example, the following rule, which does not discriminate between ships able to score one or two hits, cannot account for subjects' performance. The rule is:

> Choose the path with the least number of ships near it, providing that the number is less than five. To determine if a ship is "near" a path, draw a line from the starting position to the midpoint between the ends of the two options. Ships to the left of the line are "near" the straight path. Those to the right of the line are "near" the curved path.

Subjects using this method would have picked the correct option about 45% of the time, a much lower percentage than the 70% actually achieved. Even if subjects knew how to define "near" in a way that let them more accurately determine with which path(s) the ships should be associated, they would still not have picked the correct choice more than about 60% of the time.

Subjects based their decisions on quickly estimated ship hit sums.

Subjects reported using the "eyeball and count" method, and our data suggest that this is what they actually did. Using the average of each subject's ship hit estimates from parts IV and VII, option choices were predicted by summing over the hits from ships in each of the pictures. These predicted choices were then compared with each of the subject's actual choices in the test and retest part of the experiment (parts I and VI). In addition, in order to provide a baseline, the consistency between the test and retest choices were computed. Subjects' choices in part I of the test, when they saw each picture the first time, predicted their choices in part VI of the test, when they saw each picture a second time, 81.2% of the time. The summed ship hit estimates predicted the choices subjects made in parts I and VI of the experiment 80.5% of the time. These data thus support the hypothesis that subjects were indeed basing their decisions on the quickly estimated ship hit sums.

Discussion: hybrid decision making

In this experiment people were taught to make decisions based on a complex outcome calculation. When conditions did not let them make careful and deliberate outcome projections, they approximated the formal calculations. This approximation included estimating the hits from the individual ships. From the material discussed so far, we cannot say whether the subjects approximations integrated both outcome calculation and recognition processes. Clearly, the "count" part of the "eyeball and count" approximation is an outcome calculation. If the estimate of ship hits is based partly on recognition, then this experiment documents a hybrid of recognition and outcome calculation decision processes.

The ship hit estimates can be said to depend on recognition if people compared the positions of the hostile ships in the new barrier with the positions of ships remembered from training, and then used the remembered previously computed hits from these ships to estimate the number of hits from the new ships.

Subjects could have estimated ship hits without remembering the number of hits associated with ships seen during training. They could have instead based their estimate on the remembered lengths of their measuring tools. If they remembered these lengths, then they could estimate the number of hits from new ships by very quickly simulating in their minds the results attained by using the measuring tools. We would not consider this mode of estimating the number of hits from new ships to be an example of recognition-primed judgment or decision making.

We did collect data in this experiment suggesting that subjects relied on recognition to estimate ship hits. These were the data collected in part V of this experiment in which subjects drew contours separating the areas from which ships could score zero, one, or two hits. Rather than review these results here, we will defer this issue to the next experiment, where it is much more clearly addressed.

EXPERIMENT 2: THE INTEGRATION OF OUTCOME CALCULATION INTO RECOGNITION PROCESSES.

This experiment examined more closely the role that recognition played in the subject's ship hit estimates. It was designed to determine to what extent:

1. Subjects relied on recognition. That is, to what exent did subjects recognize ships seen during training and use the remembered number of hits from these ships in estimating the hits from new ships.

2. Subjects calculated outcomes. They could have calculated outcomes by remembering how far ships move in an hour and how far their weapons reach, and then estimating ship hits by simulating the ship movements in their minds.

3. Subjects combined recognition with outcome calculation. In this case, they would base their estimate on the remembered number of hits from previously seen ships, but use in addition information abstracted from outcome calculation to adjust these estimates when the positions of the new ships did not match exactly the postion of any of the ships seen in training.

This experiment also tested a separate conjecture:

4. Would subjects use situation features unrelated to outcome calculation if these features allowed them to reach the correct decision quickly and accurately.

Description of the task

As in the previous experiment, each subject played the role of a Battle Group commander who is facing a hostile barrier. He was presented with a situation picture (Figure 2). His Battle Group is located at the "X". The ship symbols denote the locations of hostile ships, and the irregular cross hatched shapes indicate the submarine patrol areas. The subject must choose whether to traverse the barrier along the straight path, traverse it along the curved path, or not traverse it at all, staying where he is. He was given a procedure and simple rule for his decision:

1. Estimate the numl er of hits the Battle Group will receive from the hostile forces along each path.

2. Pick the path along which you receive the fewest hits, unless this number of hits is more than six. In that case, stay at the "X."

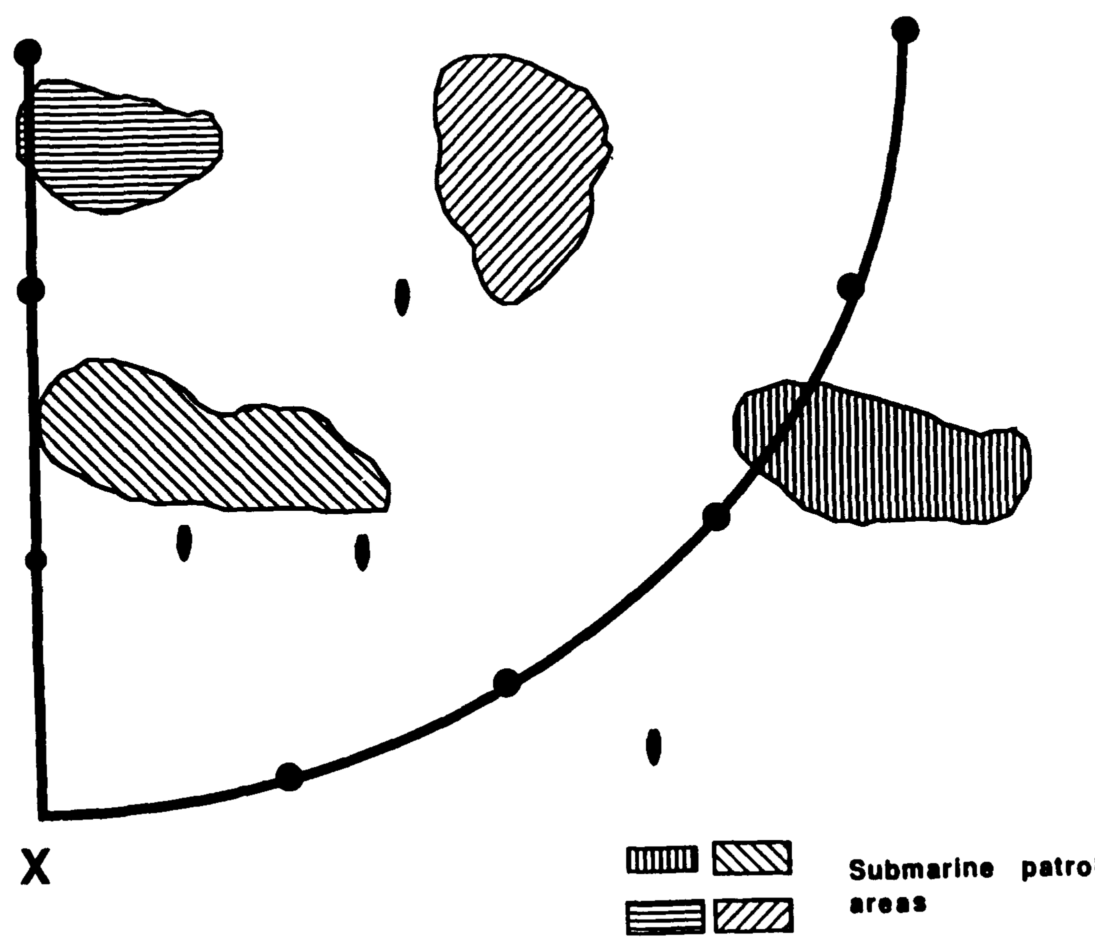


Figure 2. Sample picture of a barrier situation (experiment 2)

The subjects were provided with precise means for calculating hits from the hostile ships and submarines. They were told that their Battle Group can be attacked by enemy ships only at the black dots. There are three places they can be attacked along the straight path and five places they can be attacked along the curved path. Their Battle Group travels the distance between black dots in one hour. The Battle Group may be attacked by enemy submarines anywhere along the paths.

The computation of hits from submarines is moderately complex. It involves first dividing the area along its right-left center of mass, and then finding the center of mass for each of these halves. Each center of mass within a ship hit range scores a hit.

The method for calculating hits from the hostile ships is somewhat more complicated. The steps for this computation are:

1. Consider each path in turn.

2. When you select a path, the hostile ships will start to move into positions where they can attack. Fortunately, the hostile ships are slow. In one hour they can move only the "maximum ship movement" (shown on their measuring ruler and on Figure 3).

3. If after the first hour the ship can move to within the threat missile range of the first dot on the path, then that ship can score a hit. Determine the possible positions of enemy ships after one hour to determine the number of hits, if any.

4. After scoring one hit, a ship may change direction and try to move within the threat missile range of the next black dot along the path.

5. Estimate the positions of ships after the second hour. Those ships within the threat missile range of the second dot can score a hit.

6. Repeat the process for all successive dots along the path.

To help them with their measuring the subjects were given two paper cutouts: a circle whose radius is the threat missile range, and a ruler with divisions marked in units of maximum hourly enemy ship movement. Figure 3 shows the movements and measurements needed to calculate the hits from hostile submarines and ships. Estimating the hits from a hostile ship by projecting its future positions was intended to represent outcome calculation in decision making.

In order to encourage non-analytic estimation methods, this process was made intentionally complex. The ability of hostile ships to change direction and follow the subject's Battle Group could be confusing, for the hostile ships can sometimes score two hits by moving initially to an area between two hit spots rather than moving directly toward one of them.

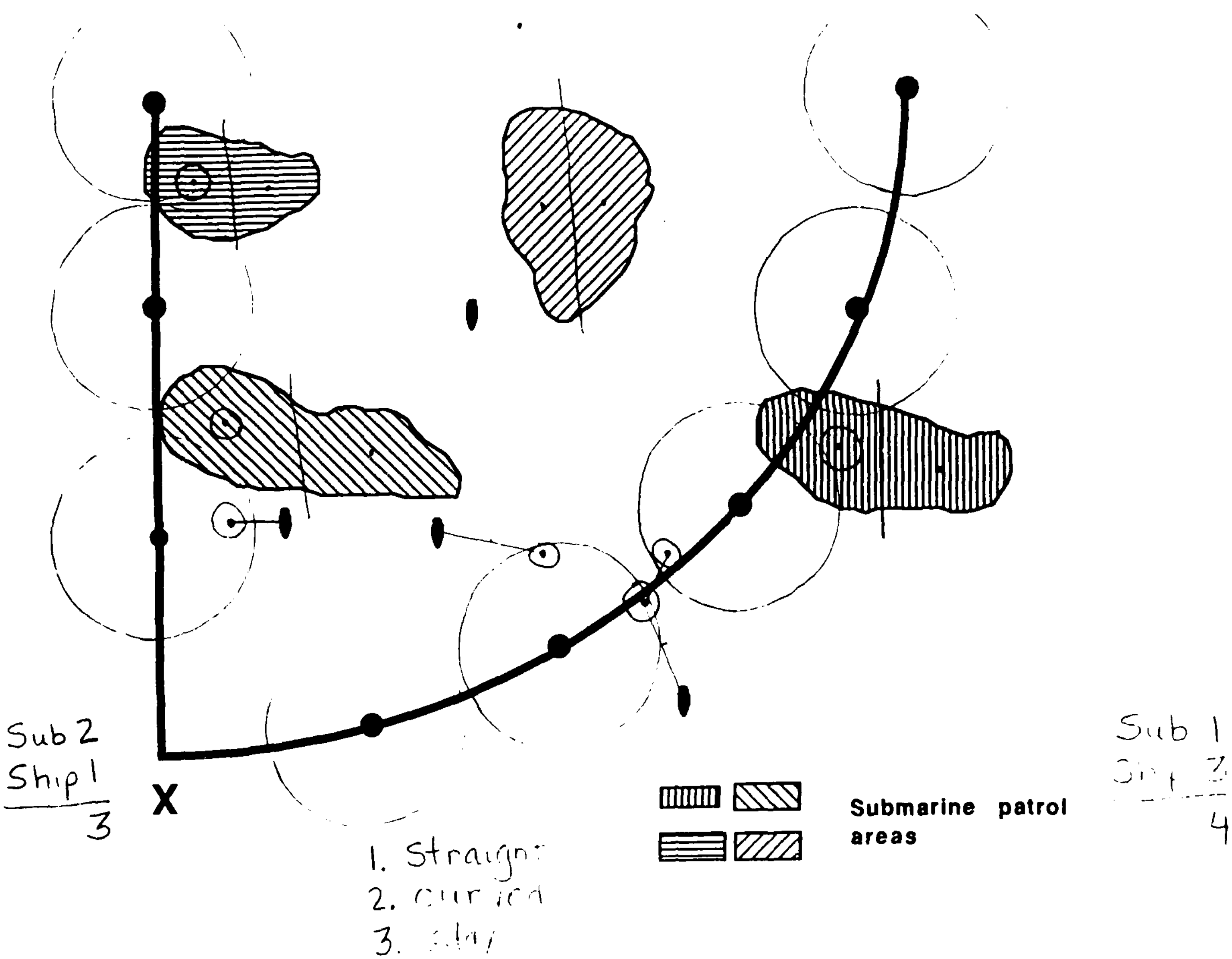


Figure 3. Barrier situation picture showing evaluation of ship and submarine hits (experiment 2).

Actually (and unknown by the subjects) it was not necessary to project the different movements of a hostile ship in order to estimate the number of hits from that ship. Instead, one could estimate these hits entirely from the initial position of the ships. There exists a set of contour range circles that mark the areas of initial positions from which hostile ships can score a hit at various hit points (the large dots along the paths). A ship whose initial position is in one of these circles can score a hit at the hit point at the center of the circle. If a ship is located in two circles, it can score two hits. If it is in three circles, however, it can still score only two hits because once a ship starts moving toward the first circle it can no longer reach the third one in time to score a hit. The subjects were **not** shown these circles, nor was there any hint given in the instructions or training that such circles exist.

In this experiment, there was a second way that observant subjects could identify the correct option without projecting the different movements of the hostile ships: they could infer the correct answer from special arbitary features in the pictures. A single ship outside the curved path (as in Figure 3) always indicated the "go straight" option is best, a pair of ships near the end of the straight path always indicated that the curved path is best, and overlapping submarine areas always indicated that the stay option is best. As in the case of the contour circles, the subjects were **not** told about these features. They were supposed to discover them by themselves during training.

During this experiment whole barriers were projected for 10 seconds, and subjects ranked the desirability of the stay, curved, straight options. Later in the experiment individual hostile ships were projected, and subjects estimated the number of hits that could be inflicted from each of these hostile ships.

Results

The subject's responses to individually projected ships revealed that ship hit estimates result from an interesting interaction between memory and outcome calculation processes.

Subjects based their estimates on recognizing previously seen ships

In this experiment, ships were projected at five different types of positions (positions A-E in Figure 4). The ships in set A were seen many times by subjects during training and testing. Ships in the pictures in sets B, C, D, and E are in locations displaced from these original locations. Ships in set B are reflections of the original 12 with respect to the paths with which they are associated. Ships in set C are displaced in a direction parallel to the ship hit circles. Ships in set D are displaced from the original location in a direction perpendicular to ship hit circles. Finally, ships in set E are displaced from ships in set D in a direction perpendicular to the ship hit circles.

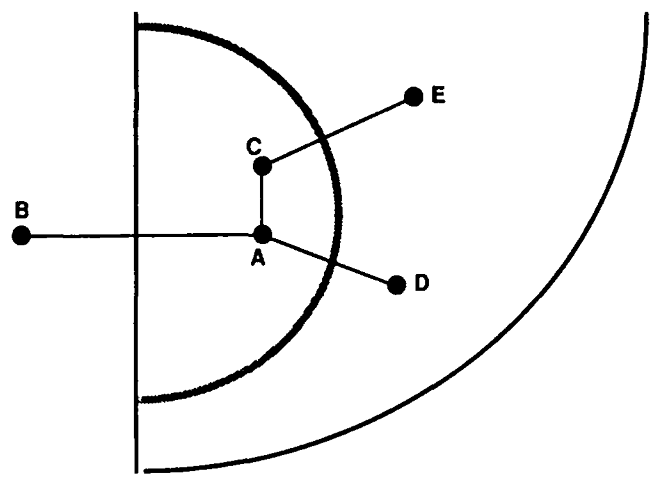
For each of these projected ships, subjects indicated those potential hit points where that ship could score a hit.

Subjects estimated the hits from previously seen ships significantly better than from ships at new positions. Overall, subjects were 78% correct for previously seen ships, and were 68% correct for ships not previously seen. For ships not seen before, the percentage correct depended how the ship was displaced from the position of previously seen ships. For example, subjects were correct only 58% of the time for ships displaced outward from the positions of previously seen ships.

Outcome calculation rules strongly impacted these estimates

For ships at new positions, subjects appeared to base their estimate of ship hits both on recognition and outcome calculation. They seemed to use the number of hits from previously seen ships as an anchor or reference, and then to adjust this number of hits by estimating the change in hits that would be caused by the displacement between the new ship and the previously seen reference ship.

SUBJECTS SHOWN SINGLE HOSTILE SHIPS, AND ESTIMATE NUMBER OF HITS ON THE BATTLE GROUP.



A. ORIGINAL - MEASURED IN 4 TRAINING PICTURES.

B. MIRROR IMAGE

C. MOVE PARALLEL TO "MENTAL RULER" CIRCLE

D. OUT FROM A.

E. OUT FROM C.

Figure 4. Different types of singly projected ships in experiment 2.

The subjects' adjustment to the number of hits reflected their understanding of how different types of displacements would affect the number of hits. For example, under the outcome calculation rules new ships whose positions are mirror reflections of previously seen ships would score exactly the same number of hits as their mirror images. The subjects estimated hits from mirror reflection ships just as accurately as they had for previously seen ships. Subjects were correct 82% of the time for these ships. Subjects were also very accurate for new ships that were the same distance from a hit point as a previously seen ship. They were 75% correct for this group.

Outcome calculation also affected subjects' estimates of new ships that were displaced radially outward on a line from the ship hit points to a previously seen ship (case D in Figure 4) . Subjects were least accurate for these ships, giving the correct answer only 58% of the time. As would be expected from the procedures for calculating ship hits, subjects estimated fewer hits for these ships than for the previously seen ship nearer to the ship hit point. Had subjects not considered the outcome calculation rules, they would have not decreased the estimate of expected number of hits.

Thus, in their ship hit estimates, subjects appeared to use both recognition and outcome calculation. They anchored their estimate on the hits from previously seen ships and then adjusted these estimates to reflect changes expected from the new ship's displacement from the previously seen ship. We can model this process as a feature-based anchor and adjustment. Subjects use features, presumably the position of the new ship relative to the straight and curved paths, to recognize previously seen ships. They also use features to assess the significance of displacements in ship position. Mirror reflections are judged not to affect the number of hits at all, and other displacements that do not change the ship's distance from a hit point are judged not to affect its ability to target that point. Changes that move a ship away from a hit point reduce the ship's ability to target that point.

Subjects did not notice or use situation features unrelated to outcome projection.

While subjects did use outcome-related features to estimate the number of ship hits, they did not notice or use "irrelevant" situation features unrelated to the outcome calculation.

All of the training materials included arbitrary features that would unambiguously denote the correct path to be chosen. During the test portion subjects looked at projected barriers and chose the best option. In the first twelve test pictures the arbitrary features still could be used to identify the correct path. In the second twelve pictures, however, subjects basing their decisions on these features would always get the wrong answer. One would expect, therefore, that subjects using these features would always get the right answer for the first twelve pictures and would never get the right answer for the second twelve.

In fact, subjects actually did better on the second set of twelve pictures than on the first set. They got 51% right in the first half (33% is expected if they were guessing) and got 69% right in the second half. Clearly, subjects did not use these special situation features which were unrelated to outcome calculation. In discussions after the test, the subjects said that they had not noticed these features.

DISCUSSION

These experiments support several general conclusions about the interplay of recognition and outcome calculation in decision making. These conclusions are:

1. Memory data able to support recognition-primed decision making will develop from experience with a decision task based on outcome calculation.

2. The memory data associate judgments with previously seen situations. In these experiments, the data are the previously seen ships and the judgments are the number of hits by a hostile ship along the paths.

3. Additional knowledge in memory enabled people to evaluate the significance of differences between a new observed situation and the remembered instances from training. This additional knowledge was the effect on "ship hits" from changes to ship position. This knowledge was directly related to the methods of calculating ship hits taught during training.

4. Subjects did not notice or use simple indicators of the best decision alternative if these indicators were not related to outcome calculation, even though such indicators could be used to quickly identify the right decision alternative.

5. The subjects' overall decision making process included components of both recognition and outcome calculation. In these experiments subjects never learned to recognize the right alternative from the overall appearance of the situation.

Although the observations reported in this paper are based on only two experiments, they emphasize what may become a central theme in decision research. As we learn more about decision making, we will no doubt discover that people draw upon a diverse set of decision making strategies. This paper focused on two of these: recognition-primed decision making and decision making based on outcome calculation. In this experiment these two modes of decision making were complementary, each contributing to different parts of the overall decision.

If in naturalistic decision making people integrate different decision strategy modes, then decision research may wish to examine the conditions under which these different processes are used, how they interact when both contribute to a decision, and how the processes affect the quality of decisions. We may wish to study the numerous task and decision maker characteristics which influence the selection and integration of these different decision making strategies.

This paper suggests another major theme for decision research: an investigation of the fundamental cognitive mechanisms underlying decision making, including processes that depend on the specific ways in which memory is organized. Here we emphasized the importance of individual remembered exemplars in recognition-primed decision making. We also suggested that knowledge of general principles of how the world works, represented here by the method for computing ship hits, augments the remembered exemplars. This general knowledge enables people to recognize the significance of differences between a current situation and a remembered previous one, and to make appropriate adjustments to the judgements and actions associated with the previous situation.

The work performed by investigators interested in cognitive mechanisms for classification may be useful in understanding recognition-primed decision making. The content and organization of memory used for recognition-primed decision making likely resembles that used in classification, for recognition-primed decision making depends on identifying features relevant to classifying a situation as "the kind of situation for which a particular type of action is likely to work."

It is not understood at present how people identify which features to use for classifying an object or event. In general these features seem to depend on the context in which the classification occurs. According to Murphy and Medin (1985), people need a "theory" to identify which features matter in classification. Such a theory would dictate to which feature people should attend, and may reflect the purpose of a category. This paper suggests that in the case of recognition-primed decision making, the "theory" is the knowledge used for outcome calculation and the features are those whose characteristics are relevant to the estimating the outcome. Features that are not related to the outcome of a decision, as in our experiment 2 or in the Lewis and Anderson (1985) experiment, tend not to be noticed or used.

As our understanding of the cognitive basis for recognition-primed decision making increases it may provide a theoretical foundation for training, planning, and decision aids that support this decision making mode. These methods could enable novices to acquire expert decision skills more rapidly than current methods do, helping novices to perceive the essential elements of a planning or decision task by enabling them to see these tasks "through the eyes of an expert."

REFERENCES

Lewis, Mathew W. and Anderson John R. (1985) Discrimination of operator schemata in problem solving: learning from examples. Cognitive Psychology, 17, 26-65.

Murphy, Gregory L and Medin, Douglas L. (1985) The role of theories in conceptual coherence. Psychological Review, 92: 289-316.

Noble, D. , Grosz, C. and Boehm-Davis, D., (1987) "Rules, schema, and decision making". Technical Report R-125-87, Engineering Research Associates, Vienna, Va.

ATTACHMENT 4

DISTRIBUTED DECISION MAKING UNDER UNCERTAINTY

Engineering Research Associates
1595 Springhill Road
Vienna, Virginia 22180

OCTOBER 15, 1989

Sponsored by

COGNITIVE AND NEURAL
SCIENCES DIVISION
Office of Naval Research
Contract N00014-84-0484

DISTRIBUTED DECISION MAKING UNDER UNCERTAINTY

ABSTRACT

These experiments evaluated how people make team decisions under uncertainty when they second guess their partner. Subjects were presented with a sequence of targets. They were told that they should shoot at light weight targets and that their partner would shoot at heavy ones. Target weight was ambiguous, but could be estimated from its size and color. An optimal decision strategy is to shoot if and only if the target's estimated weight is below a shoot/no shoot threshold. Though the subjects' behavior qualitatively resembled that expected were they following the optimal decision strategy, subjects did not seem to make their decisions this way. Rather than using a shoot/no shoot threshold, they instead based their decisions on their guesses of what their partners would do. When a partner behaved differently from normal, subjects assumed that he was estimating weight differently rather than following different rules. Subjects attempted to guess what their partners would do even for ones who shot at random.

INTRODUCTION

In distributed decision making team members working from a common plan make individual decisions to attain a common goal. Successful coordination occurs when these individual decisions support one another. Coordination errors occur when the decisions conflict.

Successful coordination depends on team members having a common understanding of the plan, an accurate understanding of the situation, and a clear idea of what other team members will do. Communications among team members can be very important to successful coordination, but communication may not be possible if communications facilities are not available or if events are occurring so fast that there is no time for communications.

Fortunately, successful coordination can still occur even when communication is not possible. For example, when a plan specifies actions to be taken by different decision makers in various kinds of situations, then coordination will be successful if all decision makers interpret the situation correctly and choose actions specified by the plan for that situation. Since in this case successful coordination depends on the team members interpreting the situation the same way, each team member when making his decision may consider how other team members are interpreting the situation.

Sometimes, coordination based on a shared understanding of a plan and a common situation interpretation can break down. This can occur if the situation is ambiguous so that some decision makers are not certain which aspects of the plan should be exercised. It can also occur if a decision maker has biases against certain kinds of actions, which he avoids if the requirement to do them under the plan is at all ambiguous.

These experiments examined decision making in two person teams in which communications was impossible and deciding what to do seems to require second guessing what what one's team partner is going to do. Data collected documented how people assessed one's team partner and predicted his choice, and how people used threat, partner, and costs of decision errors in decision making.

In the decision problem examined here, situations were represented by targets of different weights and situation assessment was represented by an estimate of a target's weight. Responsibility for shooting different types of targets was allocated between the two team members. Coordination was successful if exactly one team member shot at a target, and consequences of poor coordination were represented by penalties imposed if both team members or if neither team member shot at the target.

Since this decision problem is very simple, it is possible to compute normative decision criteria and to compare subjects actual performance with this criteria. In the computed optimal team coordination strategy, each team decision maker decides what to do by estimating the target's weight, comparing this weight with a threshold which separates different courses of action, and selecting the threshold-determined course of action. It is not necessary for the decision makers to consider explicitly his uncertainty about the target's weight, the relative penalties for different types of coordination errors, or what the other team member will do. Consideration of these factors is absorbed into the computed decision threshold.

In these experiments people's decisions qualitatively resembled those called for by the optimal strategy, but our subjects did not seem to follow the optimal process. They did not adopt a target weight threshold but instead estimated what their partners would do. Indeed, their decisions were more sensitive to their guesses of the partner's situation estimates than to their own estimates of the situation. Modeling their partner and estimating what he would do were central to their decision process, and this modeling enabled our subjects easily to accommodate partners with various types of decision biases. Interestingly, our subjects ascribed differences in a partner's behavior to differences in his situation interpretation. They did not assume that their partner was deliberately using rules which violate the common plan. This tendency to guess what partner would do was so strong that subjects continued to do this even with a partner whose behavior was completely erratic. Even with a completely "flakey" partner, subjects did not choose to ignore partner and adopt instead the mathematically superior strategy of treating the partner's actions as random.

THE DISTRIBUTED DECISION MAKING EXPERIMENT

The decision making "teams" consisted of a subject and an unseen (and nonexistent) team partner. Initially each subject was told that he would be shown pictures of targets and would be making decisions about which targets to shoot. Subjects were assigned responsibility for shooting at light targets, those weighing less the eleven pounds. They were told that target weight could be estimated from target size and darkness, and they practiced estimating target weight from its size and color until they attained a required level of proficiency. Figure 1 illustrates several of these targets.

Figure 1. Examples of targets. Weight is estimated from target size and shading.

The first part of the experiment established a baseline of individual decision making. After being trained, each subject was told that he should consider the size of the penalties for wrong decisions when deciding whether to shoot a target. One group of subjects was told that it was ten times worse to shoot a target heavier than eleven pounds than to not shoot a target lighter than eleven pounds. For the second group of subjects this ratios was reversed.

In the next part of the experiment each subject was told that he was now a member of a two person team assigned to shoot targets. Furthermore, subjects were told that because there were too many targets for one person to handle, responsibility for shooting various weight targets was divided between the subject and his partner. Subjects were assigned targets weighing less than eleven pounds, and their partners were assigned targets weighing more than eleven pounds. They were also told that different kinds of mistakes carried different penalties. For one group of subjects the penalty for neither team member shooting was ten times greater than the penalty for both shooting. For the second group of subjects, the relative magnitude of the penalties were reversed.

During testing subjects were shown a sequence of targets. For each target they were asked how much they thought the target weighed, how much they thought their partner believed it weighed, how confident partner was that the target weighed less than 11 pounds, whether they would shoot at this target, and whether partner would shoot at the target. Several times during the experiment the subjects were told that they had a new partner. For one group of subjects these partners were described as "normal", "trigger happy," "size conscious," and "flake." For the other group of subjects the partners were described as "normal,", "gun shy," and "certifiably random flake." For partners not described as normal or as a "certifiable flake", subjects were shown examples of targets which this partner shot and examples which this partner did not shoot. Subjects were not permitted to communicate with their partner during the task, who in fact did not actually exist.

NORMATIVE DECISION STRATEGY

The optimal decision strategy defines shoot/no shoot thresholds for both members of the team. These thresholds depend on the relative sizes of the penalties for various kinds of coordination errors, on the expertize of both team members, and on the apriori probability that a target of any particular weight will arrive. Once the thresholds are determined, teams will perform optimally if each team member decides whether or not to shoot solely by comparing his estimate of target weight with his shoot/ no shoot threshold. Neither team member needs to consider explicitly what his partner will do. Partner's behavior is implicit in the thresholds.

These thresholds can be readily computed when each team member independently estimates the weight of a target and when each knows the probability distribution of these estimated weights as function of the actual weight. In the discussion below we assume that the estimated weight of a target is distributed normally, with a mean equal to the target's actual weight and a variance whose inverse represents a team member's expertize at estimating target weight. We also assume that there are only two types of coordination errors: both shoot the target and neither shoots the target.

An expected loss function can be computed for each target weight given assumed shoot/no shoot thresholds for each team member and the probability distributions for estimated target weight as a function of actual target weight. This expected loss is the sum of the expected loss from both shooting and the expected loss from neither shooting. The former is the product of the penalty if both shoot and the probability that both shoot (which is the cumulative probability that the target weight estimates of both team members lie on the shoot side of their shoot/no shoot thresholds). The latter is the product of the probability that neither shoots and the penalty if neither shoots. The penalties are part of the decision environment, which in our experiment was part of the problem statement.

Optimal thresholds can be computed by finding those thresholds which minimize a global loss function. This global loss function is the sum over targets of the probability that a target of a given weight will arrive and the expected loss associated with a target of this weight.

For the assumptions specified above the thresholds so computed have the following properties:

- If the penalties for both shooting or for neither shooting are the same, then the shoot thresholds for both partners is eleven pounds.
- If the penalty for both shooting exceeds the penalty for neither shooting, then the threshold for one team member is an amount less than eleven pounds and the threshold for the other member will be this same amount in excess of eleven pounds. In order to hedge for uncertainty there is a weight "gap" where neither team member should shoot.
- The size of the weight gap depends on the expertize of the two team members, increasing monotonically with the sum of the variances on the weight estimate normal distributions. The size of the gap also increases monotonically with the ratio of penalties for both shooting and for neither shooting.
- If the penalty for neither shooting exceeds the penalty for both shooting, then the threshold pattern is reversed from that in the previous case. In order to hedge for uncertainty there is a weight "overlap" where both team members should shoot.

If the probability that one team member will shoot a particular weight target is known and fixed, then the other team member can compute his optimal decision by minimizing the global loss function as a function only of his threshold. In an extreme case of interest in these experiments, the subjects' partner fired randomly at all targets. In this case, the optimal decision strategy was either to shoot at all targets or to shoot at no targets, with the choice depending on the probability that partner would shoot and on the relative penalties for both shooting or for neither shooting.

OBSERVED DECISION BEHAVIOR

According to the analysis above, distributed decision making teams can achieve optimal performance if each team member makes his decision solely by comparing his estimate of target weight with his shoot/ no shoot threshold. In our experiments subjects did not make their decisions this way. Rather they based their decisions primarily on their guess of what their partner would do. This guess is based mostly on their estimate of their partner's estimate of target weight, and partly on the relative penalties for coordination errors. While subjects varied these estimates for different kinds of partners (normal, trigger happy, size conscious, flake) , all subjects tended to assume that their partners were like themselves and that deviations from expected shoot/ no shoot behavior reflected differences in how partners estimated target weight rather than in their criteria for shooting.

<u>Subject's decisions resemble those produced by optimal strategy</u>

Each subject was shown ten targets, one each weighing five, six, sixteen, and twenty pounds and two each weighing eight, ten, and thirteen pounds. Subjects were reasonably accurate at evaluating the weights of these targets. On the average subjects guessed that a target weighed .6 pounds than it actually did. The variance of the distribution for estimated weight as a function of actual weight was about 2.5 pounds.

Figure 2 compares the performance of the subjects with the computed optimal performance. Given the accuracy at which subjects could evaluate targets and given the penalties, subjects should have chosen to shoot four targets when the penalties discouraged shooting and to shoot eight targets when the penalties encouraged shooting.

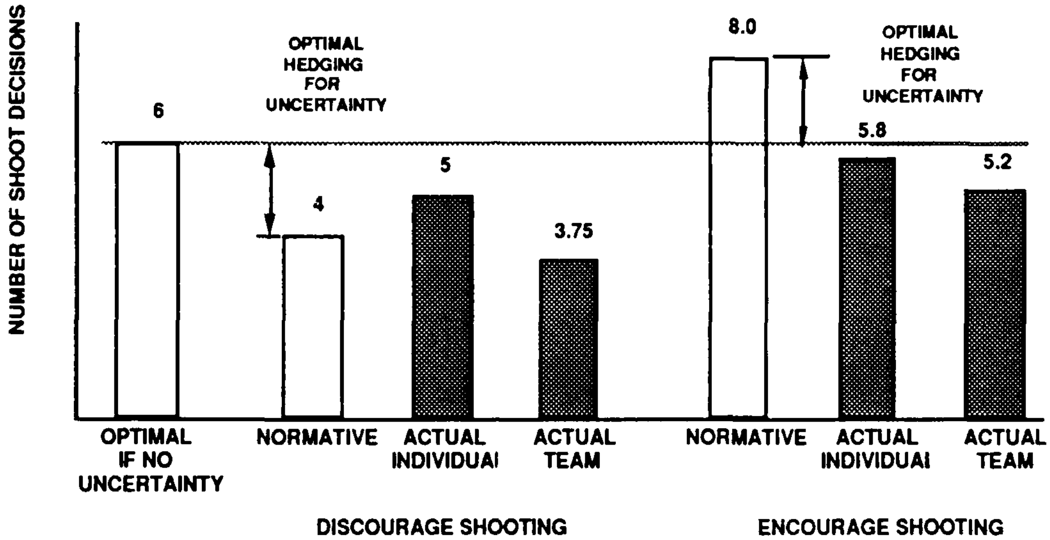


Figure 2. Subjects' shoot decision are affected by being part of a team and by penalties for decision errors.

Overall, subjects 'decisions qualitatively resembled the decisions that would be produced by the optimal strategy. Subjects hedged for uncertainty about true target weight and they hedged more with a partner than without one. This additional hedging is appropriate since adding a partner whose decisions are uncertain increases the total outcome uncertainty.

Also as expected in an optimal strategy, subjects shot more when the penalties encouraged shooting than when the penalties discouraged it. When the penalties encouraged shooting there was an overlap of several targets which subjects thought neither they nor their partner would shoot . Similarly when the penalties discouraged shooting there was a gap of several targets which subjects thought no one would shoot. Table 1 summarizes this overlap and gap. It shows the number of targets out of ten that subjects estimated that both they and their partner would shoot and the number of targets out of ten that neither they nor their partner would shoot.

<u>Subjects based shoot decisions on their estimate of partner's target weight estimate</u>

Although subjects' decisions qualitatively resembled the pattern expected were they following the optimal decision strategy, subjects did not seem to be actually making their

Type of Partner	Number of Targets (of ten) that Neither Team Member Shoots	Number of Targets (of ten) that Both Team Member Shoot
Penalty discourages shooting		
Normal	.70	--
Trigger Happy	.75	--
Penalty encourages shooting	--	.50
Normal	--	.13
Gun Shy		

Table 1. Subjects' hedge for uncertainty depends on the penalty for different types of team errors.

decisions this way. Instead of simply comparing their estimates of target weight with a shoot/no shoot threshold, they seemed to base their decision on what they thought their partners would do, which in turn depended on what they thought partner would estimate the target weight to be.

Figure 3 compares the probability that a subject will shoot as a function of his estimate of target weight with the probability of shooting as a function of his estimate of his partner's estimate. This figure summarize decisions when the subjects were told that their partners were normal and when the payoffs encouraged shooting. The pattern for the case when the payoff discouraged shooting is similar. In both cases subjects were supposed to

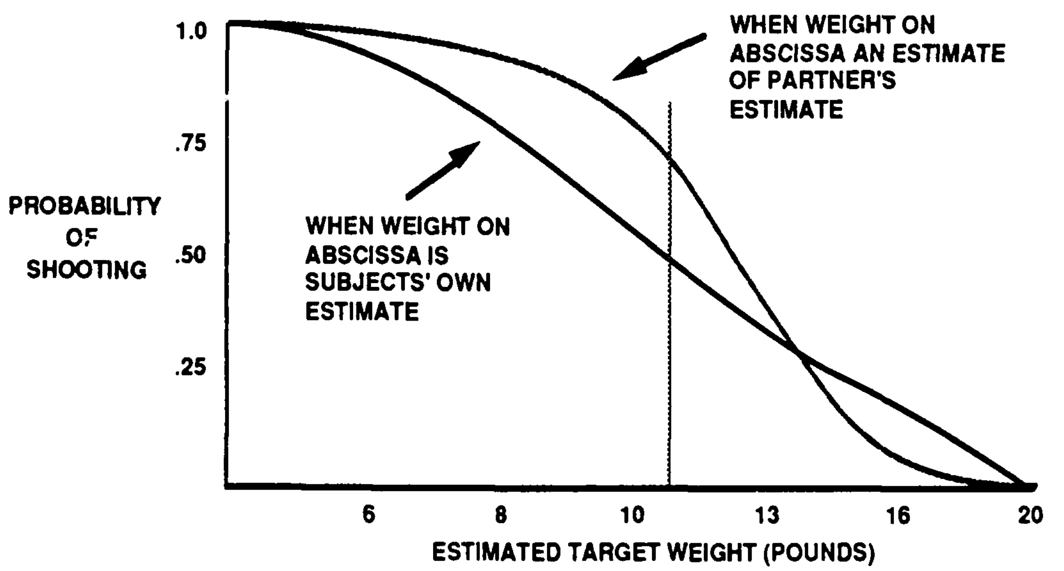


Figure 3. Pobability that a subject will shoot at a target when coordination penalties encourage shooting.

shoot at targets weighing less than eleven pounds and partner was supposed to shoot at targets weighing more than eleven pounds.

In this figure the curve of the subjects' probability of shooting is noticeably steeper when plotted as a function of his estimate of partner's estimate of target weight than when plotted as a function of his own estimate of target weight. Thus a subject's estimate of partner's weight estimate is a better predictor of his shoot/ no shoot decision than the subject's own estimate of target weight. Presumably, subjects are first estimating the target weight, then estimating partner's estimate, then deciding what partner will do, and then taking into account the penalties for different types of errors, deciding whether or not to shoot.

The importance of predicting what partner will do is supported by data relating subjects' decisions with their estimate of whether their partner will shoot . When the penalties discouraged shooting, subjects never shot when they thought that their partners would. Conversely when the penalties encouraged shooting, subjects nearly always (thirty eight in thirty nine times) shot when they thought their partners would not.

Subjects' decisions reflected their partners' characteristics

During the experiment subjects teamed with several different types of partners. For the group in which the payoff encouraged shooting these partners were "normal,", "trigger happy", "size conscious", and "flake". The partners for the group in which the payoff discouraged shooting were "normal", "gun shy", and "certifiable flake". Except for the normal partners and the "certifiable flake", subjects were introduced to these partners by a one or two sentence description and by viewing samples of the types of targets these partner's chose to shoot or not shoot. Subjects could review these examples as many times as they wanted prior to beginning the test.

Subjects' decisions depended on their partner's characteristics. Table 2 summarizes these decisions for different types of partners.

As shown in Table 3, subjects' decisions was sensitive to what they thought their partners would do. This sensitivity also held for the "gun shy" partner, who was not included in Table 2 or 3 because the expected decision changes are more complex than that for the other types. The size conscious partner considered only size when deciding whether to shoot. This partner actually shoots two large light targets which a normal partner would not shoot. On the average subjects estimated that he would shoot .875 of these targets. This partner also chooses not to shoot two small heavy targets which a normal partner would shoot. Our subjects estimated that the partner would shoot 1.125 of these.

Type of Partner	Number of Targets that should be Shot given this Partner and no Uncertainty	Number of Targets Subjects Shot
Penalty discourages shooting		
Normal	6	3.75
Trigger Happy	2	2.25
Flake	0	3.62
Penalty encourages shooting		
Normal	6	5.22
Gun Shy	8	7.78

Table 2. Subjects' decisions of when to shoot depend on their partner

As shown in Tables 2 and 3 subjects' decisions with the "flake" partner closely resembled their decisions with the "normal" partner, and their estimates of the the flake partner's decisions were very similar to their estimates of the normal partner's decisions.

Type of Partner	Number of Targets this Partner would actually Shoot if no Uncertainty	Number of Targets Subjects Estimated this Partner would Shoot
Penalty discourages shooting		
Normal	4	4.87
Trigger Happy	8	7.00
Flake	3	4.37
Penalty encourages shooting		
Normal	4	5.55
Gun Shy	2	2.33

Table 3. Subjects' estimates of the number of targets partner will shoot.

Partner's behavior attributed to situation assessment

Subjects could have attributed the decisions of the trigger happy, size conscious, gun shy and flake partners can possibly to two different causes: 1) these partners use the same decision criteria as a normal partner, but estimate target weight differently; or 2) these partners estimate weight the same as a normal partner, but use different decision criteria. In the first case, the partner is assumed to be following the common plan. Decisions that differ from those expected by the plan are caused by differences in situation assessment. In the second case, the partners assess the situation normally, but choose not to follow the common plan.

For each of these cases, subjects attributed partner's behavior mostly to biases in their estimation of target weight. One can account for the trigger happy partner's decisions by assuming that he estimates each target to be 1.75 pounds heavier than it actually is. Subjects actually assumed that this partner was adding an average of 1.7 pounds to each target's weight.

Similarly, the gun shy partners' shoot decisions can be explained by their target weight estimates. During the tests subjects first had a normal partner, who was later replaced by the gun shy partner. On the average, subjects estimated that their gun shy partner would shoot at 3.45 fewer targets than did their normal partner. Subjects also estimated that this gun shy partner would estimate that the targets weigh less than did their normal target. Of the 3.45 fewer targets fired at, changes in weight estimate account for 2.78 of the targets and "loss of nerve" account for the remaining .67 targets.

The size conscious partner based his decision only on target size and ignored its weight. Because size and weight are correlated in these experiments, most of the time subjects' estimates of their partners' actions are consistent with the partner's estimate of either the target's weight or size. For 25% of the targets presented, however, the subjects' estimates of their partners' actions were not consistent with both the partner's estimate of size and weight. In these cases, partner's weight estimate accounted for his shoot decisions 75% of the time, even though this partner actually paid attention only to target size.

People are not natural game theorists

The preceding discussion suggests that as part of their natural decision process people will predict what their partners will do and that in making these predictions people usually assume that their partners are cooperative and follow the rules when they make their decisions. Deviations from behavior expected by the rules is mostly attributed to how their partners estimate target weight rather than to their disregarding the rules.

Subjects attempted to predict what their partner would do even in the extreme case of a flakey partner who fired at random. When introduced to this partner, subjects were told that he was flakey, and were shown samples of the flake's decisions. The flake shot at three of twelve presented targets. These three targets included a six pound target medium in size and color, a ten pound small dark target, and a thirteen pound large light target. Despite seeing these examples, the subjects assumed that this flake was "not a complete flake" (otherwise he would not have been chosen for this assignment), and would be estimating target weight and making decisions sensibly. Generally, subjects attempted to understand their partners, so that they could predict what they would do. Oddly, given a partner whose behavior was so unpredictable, subjects chose to treat him as if he were normal. The distribution of target weight estimates ascribed to the flake is almost the same as the one ascribed to the normal partner, and the average number of shoot decisions by the flake was nearly the same as the number predicted for the normal partner.

Were the subjects explicitly computing and comparing the expected utilities for their shoot and don't shoot alternatives, they would have noticed that one should either shoot all the time or none of the time when given a partner who shoots at random. In this case, none of the eight subjects in our first group (some of whom were engineers) chose to adapt this optimal strategy. In order to encourage this behavior in our second group of subjects, we emphasized that the flake shot at random and we showed no examples of targets. This time two of our nine subjects adapted an optimal never shoot strategy.

Technical Reports Distribution List

Dr. Earl A. Alluisi
OUSDR (A)/R&A(E&LS)
Pentagon 3D129
Washington, D.C. 20301-3080

Dr. Michael Athans
M.I.T.
Lab for Information and
Decision Systems
Cambridge, MA 02139

Dr. Alan D. Baddeley
Director, Appl. Psychol. Unit
Medical Research Council
15 Chaucer Road
Cambridge CB2 2EF
England

Dr. Bruce Barnes
Program Director, Div. Info
National Science Foundation
1800 G Street, N.W.
Washington, D.C. 20550

Mr. Charles Bates
Director, HE Division
USAF AAMRL/HE
WPAFB, OH 45433

Mr. Normal Beck
Naval Underwater Systems Center
Combat Control Systems Dept
Code 221
Newport, RI 02840

Dr. Kenneth R. Boff
AAMRL/HF
WPAFB, OH 45433

Mr. Luiz Cabral
Naval Underwater Systems Center
Bldg. 1171/1
Newport, RI 02841

Cdr. Robert C. Carter, USN
Naval Research Laboratory
Code 5532
Washington, D.C. 20375-5000

Dr. David Castonon
Alphatech, Inc.
111 Middlesex Turnpike
Burlington, MA 01803

Dr. Marvin S. Cohen
Dec. Sci. Consortium, Inc.
1895 Preston White Drive
Suite 300
Reston, VA 22091

Dr. Stanley C. Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Defense Tech. Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314

Prof. Jay Elkerton
University of Michigan
Dept. of I&O Engineering
Center for Ergonomics
1205 Beal Avenue
Ann Arbor, MI 48109-2117

Dr. Robert A. Fleming
Naval Ocean Systems Center
Human Factors Support Group
1411 South Fern Street
Arlington, VA 22202-2896

Dr. Alfred R. Fregly
AF Office of Sci. Res.
Life Sciences Directorate
Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Sherman Gee
Office of Naval Technology
Command & Control Technol.
Code 221
800 N. Quincy Street
Arlington, VA 22217-5000

Jeffrey Grossman
Naval Ocean Systems Center
Code 4403
Bldg. 334
San Diego, CA 92152-6800

Dr. Thomas E. Hanna
Naval Sub. Med. Res. Lab.
Behavioral Sciences Dept
Naval Submarine Base
Groton, CT 06349-5900

Mr. Paul Heckman
Naval Ocean Systems Center
San Diego, CA 92152

Prof. Robin Hogarth
University of Chicago
Grad. School of Business
1101 E. 58th Street
Chicago, IL 60637

Prof. James H. Howard, Jr.
Department of Psychology
Human Performance Lab.
Catholic University
Washington, D.C. 20064

Prof. Douglas E. Hunter
Defense Intelligence College
Washington, D.C. 20374

Dr. Bonnie E. John
Dept of Computer Science
Carnegie-Mellon University
Pittsburg, PA 15213

Dr. Edgar M. Johnson
Technical Director
US Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. William B. Johnson
Search Technology, Inc.
4725 Peachtree Corners Circle
Suite 200
Norcross, GA 30092

Capt. T.E. Jones, MSC, USN
Aviation Medicine & Human
Performance (Code 404)
Naval Medical R&D Cmd.
National Capital Region
Bethesda, MD 21814-5044

Dr. Michael Kaplan
Director, Office Basic Res.
U.S. Army Research Institute
Alexandria, VA 22333-5600

Dr. Milton S. Katz
U.S. Army Research Institute
ARIESCO
Box 65
FPO, NY 09510-1500

Prof. David L. Kleinman
Electrical Engineering &
Computer Science Dept.
University of Connecticut
Storrs, CT 06268

Dr. Arthur F. Kramer
Department of Psychology
University of Illinois
Champaign, IL 61820

Mr. Keith Kramer
Naval Research Laboratory
Code 5532
Washington, D.C. 20375-5000

Prof. John Lehoczky
Department of Statistics
Carnegie-Mellon University
Pittsburgh, PA 15213

Mr. I. Lemaire
Naval Ocean Systems Center
Code 09
Engrg. & Computer Science
San Diego, CA 92152

Dr. Alexander H. Levis
Laboratory for Information
and Decision Systems
M.I.T.
Cambridge, MA 02139

Dr. Michael Letsky
Office of the Chief of
Naval Operations
OP-01B7
Washington, D.C. 20350

Office of the Chief of
Naval Operations
OP-933D3
Washington, D.C. 20350-2000

Office of Naval Research
Special Asst. for Marine
Corps Matters
Code 00MC
800 N. Quincy Street
Arlington, VA 22217-5000

Office of Naval Research
Perceptual Science Program
Code 1142PS
800 N Quincy Street
Arlington, VA 22217-5000

Dr. Judith Orosano
Research Psychologist
ARI
5001 Eisenhower
Alexandria, VA 22333

Dr. Jesse Orlansky
Inst. for Defense Analyses
1801 N. Beauregard Street
Alexandria, VA 22311

Lt. Dennis McBride, USN
Pacific Missile Test Center
Human Factors Branch
Pt. Mugu, CA 93042

Prof. John W. Payne
Fuqua School of Business
Ctr. for Decision Studies
Duke University
Durham, NC 27706

Dr. Richard Pew
BBN Laboratories, Inc.
Experimental Psychol. Dept.
70 Fawcett Street
Cambridge, MA 02238

Dr. W. A. Rizzo
Head, Human Factors, Div.
Naval Training Systems Ctr.
12350 Research Parkway
Orlando, FL 32826-3224

Prof. Scott Robertson
Department of Psychology
Rutgers University
Busch Campus
New Brunswick, NJ 08903

Dr. William B. Rouse
Search Technology, Inc.
4725 Peachtree Corners Circle
Suite 200
Norcross, GA 30092

Dr. Eduardo Salas
Naval Training Systems Center
Code 712
Orlando, FL 32813

Dr. Walter Schneider
University of Pittsburgh
LRDC Bldg.
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. James Shantear
Decision Research and
Managment Science
NSF
1800 G. Street, N.W.
Washington, D.C. 20550

Dr. Randall P. Shumaker
Naval Research Laboratory
Code 5510
U.S. Navy Center for ARAI
Washington, D.C. 20375-5000

Dr. H. Wallace Sinaiko
Manpower Research and
Advisory Services
Smithsonian Institution
801 N. Pitt Street, Suite 120
Alexandria, VA 22314-1713

LCdr. Timothy Singer, USN
Naval Air Development Center
Human Factors Engineering Div.
Warminster, PA 18974

Dr. Paul Slovic
Decision Research, Inc.
1201 Oak Street
Eugene, OR 97401

Dr. Dana L. Small
Naval Ocean Systems Center
Code 33
San Diego, CA 92152-5000

Mr. James G. Smith
Office of Naval Research
Code 121
800 N. Quincy Street
Arlington, VA 22217-5000

Mr. Yale Smith
Rome Air Development Center
RADC/COAD
Griffis AFB, NY 13441-5700

Prof. Harry L. Snyder
Dept. of Industrial Engineering
VPISU
Blacksburg, VA 24061

Capt. Frank Snyder, USN
Naval War College
Operations Department
Newport, RI 02841

Cdr. S. Snyder
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Prof. Michael Sovereign
Naval Postgraduate School
Code 74
Joint C3 Curriculum
Monterey, CA 93943

Mr. Steven L. Speidel
Naval Ocean Systems Center
Code 632
San Diego, CA 92152-5000

Dr. John Tangney
AF Office of Sci. Res.
Life Sciences Directorate
Bldg. 410
Bolling AFB, D.C. 20332-6448

Prof. Demosthenis Teneketzis
Dept. of Electrical Engineering
and Computer Science
University of Michigan
Ann Arbor, MI 48109-1109

Dr. Martin A. Tolcott
Decision Science Consortium
1895 Preston White Drive
Suite 300
Reston, VA 22091-4369

Dr. Douglas Towne
U. of Southern California
Behavioral Technology Lab.
1845 South Elena Avenue
Fourth Floor
Redondo Beach, CA 90277

Dr. Amos Tversky
Stanford University
Department of Psychology
Bldg. 420
Stanford, CA 94305

Prof. William R. Uttal
Department of Psychology
Arizona State University
Tempe, AZ 85287-1104

Dr. Harold P. Van Cott
Committee on Human Factors
NAS-National Res. Council
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Dr. John Weisz
Technical Director
U.S. Army Human Engineering
Laboratory
Aberdeen P.G., MD 21005

Dr. Elizabeth M. Wenzel
NASA-Ames Research Center
MS #239-3
Moffett Field, CA 94035

Dr. Robert Wherry
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

Prof. Christopher D. Wickens
Department of Psychology
University of Illinois
Urbana, IL 61801

Dr. Diane Williams
Naval Ocean Systems Center
Code 4403
Bldg. 334
San Diego, CA 92152-6800

Dr. Wayne Zachary
CHI Systems, Inc.
Gwynedd Plaza III
Bethlehem Pike
Spring House, PA 19477